



**Consortio CSIC-IRTA**  
**Consejo Superior de Investigaciones Científicas**  
Laboratorio de Genética Molecular Vegetal  
Instituto de Biología Molecular de Barcelona  
Departamento de Genética Molecular

**FUNDACIÓN  
PRESIDENTE  
ALLENDE**



**Universidad Autónoma de Barcelona**  
Facultad de Ciencias  
Departamento de Bioquímica y Biología Molecular  
Área de Bioquímica y Biología Molecular

# **GENES IMPLICADOS EN EL DESARROLLO DE LA SEMILLA DE *Arabidopsis thaliana* (L.): CARACTERIZACIÓN DE LOS GENES *AtAnkTm***

CRISTIAN MARCELO del CARMEN BECERRA BAEZA

Barcelona, Marzo de 2006



**Universidad Autónoma de Barcelona**  
Facultad de Ciencias  
Departamento de Bioquímica y Biología Molecular  
Área de Bioquímica y Biología Molecular

**GENES IMPLICADOS EN EL  
DESARROLLO DE LA SEMILLA DE  
*Arabidopsis thaliana* (L.):  
CARACTERIZACIÓN DE LOS  
GENES *AtAnkTm***

Memoria de tesis doctoral presentada por Cristian Marcelo del Carmen Becerra Baeza para optar al grado de Doctor en Biotecnología por la Universidad Autónoma de Barcelona.

Este trabajo se ha realizado en el Departamento de Genética Molecular del Instituto de Biología Molecular de Barcelona (CSIC), bajo la dirección de los Doctores PERE PUIGDOMÈNECH ROSELL y CARLOS M. VICIENT SÁNCHEZ y la tutoría de la Dra. MARÍA CARMEN MARTÍNEZ GÓMEZ

Con el consentimiento de,

Director de Tesis

Director de Tesis

Tutor

Autor

Dr. Pere Puigdomènech R.    Dr. Carlos Vicient S.    Dra. M. Carmen Martínez G.    Cristian Becerra Baeza

*A mi Querida Violeta,  
a tu vida llena de sacrificios,  
a tu sabiduría,  
a tu amor de madre.*

“... y la vida en sí es solo eso, un conjunto de pequeños detalles que en armonía o no, van definiendo nuestro destino...”

Cristian Becerra B.

## AGRADECIMIENTOS

Este trabajo ha sido posible gracias a la ayuda y el apoyo de muchas personas, por ello quiero agradecer...

En primer lugar al Dr. Pere Puigdomènech, por el apoyo y la confianza que depositó en mi, su continua preocupación por mi trabajo y las enseñanzas que poco a poco fue traspasándome en nuestras conversaciones.

Muy especialmente al Dr. Carlos M. Vicient. Bueno, si hay alguien a quien debo dar las gracias por hacer posible finalizar esta memoria de tesis es a este señor. De él he aprendido mucho, tanto en la parte profesional como en la humana. Trabajador incansable que te contagia con su dedicación por la ciencia y su continuo ánimo en pro de no bajar la guardia y seguir adelante frente a todos los contratiempos (que en esta área son muchos) que se fueron presentando en este largo camino. Por todas sus enseñanzas, su gran valor como persona, su especial buen sentido del humor y por haberme aguantado tanto tiempo...¡ Muchas Gracias!!!

A don Aldo González B., por su constante preocupación en mi formación humana y profesional y por ser uno de los artífices en que me dedicase a la investigación. Querido Aldo, muchas gracias!.

A la Dra. María Carmen Martínez, por admitirme bajo su tutela en la Universidad y su constante buena disposición para atenderme y ayudarme.

A Núria, mi querida gran amiga y compañera. Son muchos los momentos inolvidables compartidos, buenos y malos. Gracias por estar siempre apoyándome y aguantándome (que eso sí que es difícil) durante tanto tiempo y muy especialmente, este último período, uno de los más complejos de esto que llamamos vida. Mis mejores deseos para tu trabajo, seguro que serás una gran investigadora!.

A Ignacio, con quien aprendí a dar los primeros pasos en este mundo de la genética molecular, por su apoyo y preocupación por mi trabajo, por compartir tantas penas y alegrías y muy especialmente por su calidez humana, Gracias Ignacio!.

A mis ex compañeros de laboratorio, Sergi F., con quien compartí momentos muy gratos durante los primeros años y por nuestra innovación en el área médica que espero algún día podamos patentar: la puntapuntura. A Víctor (que ahora está en el exilio en el lab. Naranja) un verdadero libro abierto con el que hablar se transforma en una aventura de mundos desconocidos y personajes inimaginables.

A todos mis amigos y compañeros del Departamento de Genética Molecular del IBMB (si dejo a alguien fuera, mil perdones).

A Jordi Q. y Laura R., por su amistad, apoyo y preocupación. Son muchos los momentos inolvidables. Espero nos veamos en Chile!!!

A las Wendys, Elizabeth e Irma, con las que he compartido mucho buenos momentos tanto dentro como fuera del trabajo. Bueno, seguiré esperando mi carnet de Wendísimo para poder formar parte de vuestro exclusivo equipo...hasta siempre!

A Ana Paula y Pedro, quienes me han dado su cariño y apoyo y me han recibido en casa como uno más de su familia. Gracias por permitirme disfrutar tan gratos momentos.

A Cristina y Valeria, por sus siempre alegres compañías y sus coreografías en la sala de espectáculos del gran lab. Verde.

A David Caparrós, por su amistad, sus consejos y apoyo y sus bromas. Por compartir tan gratos momentos junto a nuestro glorioso equipo de fútbol (IBMB, Ciencia y ley) y por las carrerillas en los pasillos. Gracias David!

A Mercè, quien siempre me ha estado dando ánimos para poder seguir y terminar este trabajo. A Hédia, quien llevo hace poquito, pero que también me ha animado a terminar esta memoria.

A mi amiga Marta, por los inolvidables momentos compartidos en el trabajo y fuera de él. Gracias por tu constante apoyo y preocupación. Espero que vayas a Chile, si es que cumples con tu palabra...

A la Dra. Martine Devic y Jocelyne Guillemot de la Universidad de Perpignan por su disposición y ayuda prestada en la técnica de hibridación *in situ*. A don Juan Martín Villodre de la Fac. de Farmacia (U. de Barcelona) por sus consejos en el tratamiento del polen.

A todos mis compañeros y amigos que quedan y algunos que ya se han ido, me gustaría dedicarles a todos unas líneas, pero he de ser breve...

A mi compañero Sami, quien además ha tenido que aguantarme en casa... a Alicia, Javi. A la Dra. Victoria Lumbreras. Carles, Beatriz, Bía, Sonia, Jorge, Lidia. A Elisenda, Luis, Céline Sorin, Jaume, Desi, Mireia. A Mariana Rodríguez, Fathy Sonbol, André, Alex, Enríque, Patricia. A Nora, José Luis, Inma C., a los Drs. Miquel Vendrell y Josep Torné. A los cucas, Oscar, Nuria, Dani, Vivi y Paula. A mi amigo Roberto. A Maite por toda su buena disposición siempre y su aprecio. A Mina. A Néstor, Enric, Pep, Céline Loot, Nahuel, Juanjo, Mariano, Juan, Paula, Soraya, Cristina y Maida. A Pilar Fontanet, Maricarmen, Leire, César, a Montse, Eva. Raúl, A todos...

A mis amigos de fuera: Lily y Rodrigo, Marina, Nelly y Manuel, Antonieta y Carlos, Sole.

Muy especialmente a Renza Salazar B., por su cariño y su paciencia, ya ha pasado mucho tiempo, no?. Ya vuelvo...Gracias por todo tu apoyo siempre, a pesar de la distancia y, por estar conmigo en los momentos más difíciles.

A mis viejos con cariño, Alamiro Salas (Q.E.P.D.) y J. Aurelio Contreras.

Este período de desarrollo de tesis ha podido realizarse gracias a una beca predoctoral de la Universidad Autónoma de Barcelona y la Fundación Presidente Allende. También debo mi agradecimiento al Consorcio CSIC-IRTA.

## ABREVIATURAS

$\alpha$	alpha
$\beta$	beta
$\kappa$	kappa
$\lambda$	lambda
aa	Aminoácido
ABA	Ábscisic acid (ácido abscísico)
ABRC	Arabidopsis Biological Resource Center
AGI	Arabidopsis Gene Index
ANK	Anquirina
Arabidopsis	<i>Arabidopsis thaliana</i>
AtAnkTm	<i>Arabidopsis thaliana</i> Ankyrin Transmembrane
ATISLA	<i>Arabidopsis thaliana</i> Immature Seed Library
atm	Atmósfera(s)
BCIP	5-Bromo-4-Choro-3-Indolyl Phosphate
Br Et	Bromuro de Etidio
BSA	Bovine Serum Albumin (albúmina sérica bovina)
°C	Celsius degrees (grados centígrados)
cDNA	DNA complementario
cm	centímetro(s)
Col-0	Columbia 0
DAPI	4,6-Diamidino-2-fenilindol
dbEST	Base de datos de ESTs
DDA	Días después de antesis
DEPC	DiEtilPiroCarbonato
DIG	Digoxigenina
DNA	DeoxyriboNucleic Acid (ácido desoxirribonucleico)
DO	Densidad Óptica
dNTP	deoxyNucleotide Tri-Phosphate (desoxinucleotido tri-fosfato)
DTT	DitioTeTriol
EDTA	Ácido Etileno Diamino Tetracético
EST	Expressed Sequence Tag (secuencia expresada, transcrita)
EtOH	Etanol
g	gramo(s)
g	Unidad de aceleración
gDNA	DNA genómico
GEV	Gene Expression Visualization
GFP	Green Fluorescent Protein
GO	Gene Ontology
h	Horas
H <sub>2</sub> O	agua
H <sub>2</sub> O <sub>dd</sub>	Agua doblemente destilada
HPLC	Cromatografía líquida de alta presión
IPTG	Isopropil- $\beta$ -D-tiogalactopiranosido
KAc	Kalium Acetate (acetato de potasio)
Kb, kb	Kilobase(s)
kDa	KiloDaltons
kg	Kilogramo(s)
kJ	Kilojulio
l	Litro
LB	Luria-Bertani
M	Molar
Mb	Megabase
$\mu$ g	microgramo(s)
$\mu$ l	Microlitro
$\mu$ M	microMolar

min	minuto(s)
mM	miliMolar
mm	milimetro(s)
MOPS	3-(N-Morpholino)propanesulfonic acid
MPM	Marcador de Peso Molecular
mRNA	RNA mensajero
NaAc	Acetato de sodio
NASC	The Nottingham Arabidopsis Stock Centre
NBT	Nitro blue tetrazolium
NCBI	Nacional Centre for Biotechnology Information
ng	nanogramo(s)
nm	nanómetro(s)
ORF	Open Reading Frame (pauta de lectura abierta)
pb	pares de bases
PCR	Polymerase Chain Reation (reacción en cadena de la polimerasa)
pI	Punto isoeléctrico
pmol	Picomol(s)
RNA	RiboNucleic Acid (ácido ribonucleico)
rpm	revoluciones por minuto
RT	Transcriptasa reversa
RT-PCR	Reverse Transcription – Polymerase Chain Reaction
s	segundo(s)
SAIL	Syngenta Arabidopsis Insertion Lines
SDS	Sodium Dodecyl Sulfate
SIGnAL	The Salk Institute Genome Analysis Laboratory
SMART	Simple Modular Architecture Research Tool
SOTA	Self-Organising Tree Algorithm
SSC	Citrato de sodio salino
ssDNA	DNA de cadena simple
Ta	Temperatura ambiente
TAE	Tris-acetato EDTA
TAIR	The Arabidopsis Information Resource
Taq	<i>Thermus aquaticus</i>
T-DNA	Transfer DNA
TIGR	The Institute for Genomic Research
Tm	Temperatura de fusión
TM	Transmembrana
TMEV	TIGR Multiple Experiment Viewer
U	Unidad(es)
UTR	UnsTranslated Region (Región no traducida)
UV	Ultra Violeta
V	Volumen
w/v	weight/volume (peso/volumen)
X-Gal	5-bromo-4-cloro-3-indotil-β-D-galactopiranósido



## INDICE

	<u>Pág.</u>
<b>INTRODUCCIÓN</b>	1
Capítulo I <i>Arabidopsis thaliana</i>	3
I.1 <i>Arabidopsis thaliana</i> como sistema modelo del desarrollo vegetal	3
I.2 Estructura corporal y ciclo de vida	4
I.3 El genoma de <i>Arabidopsis thaliana</i>	6
I.4 Herramientas informáticas para <i>Arabidopsis thaliana</i>	9
I.5 Desarrollo del polen en <i>Arabidopsis thaliana</i> (microsporogénesis)	10
I.6 Desarrollo embrionario en <i>Arabidopsis thaliana</i>	12
I.6.1 Formación del patrón: etapa morfogénica	13
I.6.2 Maduración y Latencia	15
I.6.3 Control genético de la embriogénesis	15
Capítulo II Dominios Proteicos	19
II.1 Proteínas con dominios transmembrana	19
II.2 Repeticiones anquirina	20
II.3 Proteínas de plantas con repeticiones anquirina	23
<b>OBJETIVOS</b>	25
<b>RESULTADOS</b>	29
Capítulo I Identificación de genes que se expresan específicamente durante el desarrollo temprano de la semilla de <i>Arabidopsis thaliana</i>	31
I.1 Secuenciación de ESTs de semillas inmaduras de <i>Arabidopsis thaliana</i>	34
I.2 Selección <i>in silico</i> de genes que se expresan específicamente en semillas inmaduras	36
I.3 Validación experimental de los patrones de expresión de genes seleccionados	46
I.4 Clasificación funcional de los genes seleccionados	48
I.5 Patrones de expresión de genes durante el desarrollo de semilla y silicua	50
I.6 Redundancia genética y fenotipos mutantes	52
Capítulo II Genes que codifican proteínas con repeticiones anquirina y dominios transmembrana implicados en la embriogénesis de <i>Arabidopsis thaliana</i>	53
II.1 Repeticiones anquirina en <i>Arabidopsis</i>	56
II.1.1 Secuencia consenso de las repeticiones anquirina en <i>Arabidopsis</i>	58
II.1.2 Proteínas que contienen repeticiones anquirina en <i>Arabidopsis</i>	59
II.1.2.1 Proteínas con sólo repeticiones anquirina	63
II.1.2.2 Grupo con dominio BTB	63
II.1.2.3 Proteínas quinasas	64
II.1.2.4 Proteínas con dedos de zinc	64

	<u>Pág.</u>
II.1.2.5 Canales de potasio	64
II.1.2.6 Proteínas <i>Ring finger</i>	65
II.1.2.7 Proteínas con dominios de activación ARF GTPasa	65
II.1.2.8 Proteínas que contienen motivos de unión a calmodulina	65
II.1.2.9 Proteínas de unión a Acil-CoA	66
II.1.2.10 Proteína con cromodominio	66
II.1.2.11 Helicasa	66
II.1.2.12 Otras proteínas	66
II.2 Genes que codifican proteínas con repeticiones anquirina y dominios transmembrana en Arabidopsis	67
II.2.1 Análisis filogenético	67
II.2.1.1 Genes que codifican proteínas ANKTM en Arabidopsis	67
II.2.1.2 Genes <i>AnkTm</i> en otras especies vegetales	69
II.2.2 Organización genómica de los genes <i>AtAnkTm</i>	72
II.2.2.1 Distribución de exones e intrones	72
II.2.2.2 Distribución cromosómica	72
II.2.3 Expresión de los genes <i>AtAnkTm</i>	75
II.2.3.1 Genes control	78
II.2.3.2 Genes de la familia I	80
II.2.3.3 Genes de la familia II	85
II.2.3.4 Genes de la familia III	89
II.2.3.5 Genes de la familia IV	91
II.2.3.6 Genes de la familia V	95
II.2.3.7 Genes de la familia VI	97
II.2.4 Mutantes letales de la familia I	98
II.2.5 Análisis del gen <i>AtAnkTm28</i> de la familia IV	102
II.2.5.1 Representación esquemática del gen <i>AtAnkTm28</i>	102
II.2.5.2 Patrón de expresión del gen <i>AtAnkTm28</i>	103
II.2.5.3 Patrón espacial de expresión del gen <i>AtAnkTm28</i>	105
II.2.5.4 Localización subcelular de la proteína ATANKTM28	106
<b>DISCUSIÓN</b>	109
Capítulo I	111
I.1 Obtención de nuevas secuencias de ESTs	111
I.2 Identificación de genes de expresión específica de semilla inmadura	113
Capítulo II	121
II.1 Repeticiones anquirina en Arabidopsis	121
II.2 Genes <i>AtAnkTm</i> de <i>Arabidopsis thaliana</i>	123
II.3 Patrones de expresión de los genes <i>AtAnkTm</i> de <i>Arabidopsis thaliana</i>	124
II.4 Análisis de líneas mutantes de inserción de T-DNA	127
II.5 Posibles funciones de los genes <i>AtAnkTm</i>	128

	<u>Pág.</u>
II.6 Estudio del gen <i>AtAnkTm28</i> durante la embriogénesis de <i>Arabidopsis thaliana</i>	132
<b>CONCLUSIONES</b>	135
<b>MATERIALES y MÉTODOS</b>	139
I. Materiales	141
1.1 Material Vegetal	141
1.1.1 Especies empleadas y condiciones de crecimiento	141
1.1.2 Mutantes	141
1.1.3 Esterilización de semillas de <i>Arabidopsis thaliana</i>	142
1.1.4 Cultivo <i>in vitro</i>	142
1.2 Cepas bacterianas y vectores de clonaje	142
II. Métodos	143
2.1 Extracción de ácidos nucleicos	143
2.1.1 Extracción de DNA plasmídico	143
2.1.1.1 Minipreparaciones de DNA plasmídico	143
2.1.1.2 Preparaciones a gran escala	143
2.1.2 Extracción de DNA genómico	144
2.1.3 Extracción de RNA	144
2.1.3.1 Método del LiCl	144
2.1.3.2 Método del Trizol	145
2.1.3.3 Tratamiento del RNA con DNAsa	145
2.2 Electroforesis de ácidos nucleicos	146
2.2.1 Electroforesis de DNA en gel de agarosa	146
2.2.2 Electroforesis de RNA en gel de formaldehído/agarosa	146
2.2.3 Recuperación de DNA a partir de gel de agarosa	147
2.3 Modificaciones generales del DNA	147
2.3.1 Digestión con enzimas de restricción	147
2.4 Subclonaje de fragmentos de DNA en plásmidos	148
2.4.1 Ligación de fragmentos de DNA a un vector de clonación	148
2.4.2 Preparación de células competentes	148
2.4.2.1 Método del CaCl <sub>2</sub>	148
2.4.2.2 Para electroporación	149
2.4.3 Transformación de células competentes de <i>E. coli</i>	149
2.4.3.1 Transformación por choque térmico	149
2.4.3.2 Transformación por electroporación	150
2.5 Secuenciación	150
2.6 Genoteca de cDNA	151
2.6.1 Síntesis de la primera cadena de cDNA	151
2.6.2 Síntesis de cDNA de doble cadena	151
2.6.3 Ligación de cDNA	151
2.7 Reacción en cadena de la polimerasa (PCR)	151
2.7.1 Observaciones generales	151
2.7.2 RT-PCR semicuantitativa	152
2.7.3 Oligonucleótidos	153
2.8 Transferencia e hibridación de ácidos nucleicos	156
2.8.1 Transferencia de ácidos nucleicos	156

	<u>Pág.</u>
2.8.1.1 Transferencia para Southern	156
2.8.1.2 Transferencia para northern	157
2.8.2 Marcaje y purificación de sonda	157
2.8.2.1 Marcaje	157
2.8.2.2 Purificación	158
2.8.3 Hibridación	158
2.8.4 Deshibridación	159
2.9 Técnicas de detección <i>in situ</i>	159
2.9.1 Fijación e inclusión en parafina	159
2.9.1.1 Fijación	159
2.9.1.2 Inclusión	159
2.9.1.3 Preparación de los portaobjetos	160
2.9.1.4 Desparafinación, deshidratación y permeabilización	160
2.9.2 Síntesis de ribosondas	160
2.9.2.1 Linealización	161
2.9.2.2 Reacción de transcripción	161
2.9.2.3 Cuantificación de las sondas	161
2.9.3 Hibridación	162
2.9.4 Inmunodetección y revelado	162
2.10 Tinciones histológicas	162
2.10.1 Tinción DAPI	162
2.10.2 Tinción Naranja de Acridina	163
2.10.3 Tinción Azul de Anilina	163
2.11 Transferencia de DNA mediante microbombardeo	163
2.11.1 Preparación de los microproyectiles	164
2.11.2 Precipitación del DNA	164
2.11.3 Preparación PDS1000/He y Bombardeo	164
2.12 Microscopía Electrónica de Barrido	165
2.13 Análisis <i>in silico</i>	165
<b>BIBLIOGRAFÍA</b>	167
<b>APÉNDICE I</b>	
Computational and experimental analysis identifies Arabidopsis genes specifically expressed during early seed development	
<b>APÉNDICE II</b>	
Genes seleccionados por sustracción de ESTs	
<b>APÉNDICE III</b>	
Ankyrin repeat-containing proteins in Arabidopsis: characterization of a novel and abundant group of genes coding ankyrin-transmembrane proteins	

## **INTRODUCCIÓN**

---

## Capítulo I. *Arabidopsis thaliana*

### I.1 *Arabidopsis thaliana* como sistema modelo del desarrollo vegetal

Se denomina sistema modelo a cualquiera de los organismos experimentales en cuyo estudio se concentran los esfuerzos de un grupo amplio de equipos de investigación, con el fin de obtener conclusiones que puedan ser aplicables a otras especies (Bolker, 1995). Los criterios para elegir una especie como modelo son diversos, siendo características comunes a la mayoría de ellos la de tratarse de organismos de pequeño tamaño, de ciclo de vida corto y de mantenimiento simple y económico en el laboratorio.

Algunos organismos que han sido utilizados como modelo son *Escherichia coli* y sus fagos, que sentaron las bases de la biología molecular, *Drosophila melanogaster* (Leptin, 1994) y *Caenorhabditis elegans* (Hope, 1994), que sentaron las bases del desarrollo animal o *Xenopus* (Slack, 1994), que sirvió para establecer las bases del desarrollo embrionario. La genética clásica estableció el maíz (*Zea mays*) como planta modelo, pero la genética molecular ha establecido como principal planta modelo a *Arabidopsis thaliana* (Meyerowitz, 1994).

*Arabidopsis thaliana* posee una serie de características que hacen de ella un excelente especie modelo. Es una planta de ciclo de vida corto (6 semanas cultivada a 25 °C bajo iluminación continua), de pequeño tamaño (unos 30 cm de altura), muy prolífica (capaz de producir hasta 10.000 semillas por planta), autógena, posee un genoma muy pequeño (unas 157 Mb; Bennett *et al.*, 2003) repartido en cinco cromosomas que han sido secuenciados (The Arabidopsis Genome Initiative, 2000), y que se puede transformar de manera eficiente utilizando *Agrobacterium tumefaciens*. A pesar de las pequeñas dimensiones de su genoma haploide presenta las características típicas de otras angiospermas en lo referente a morfología, anatomía, crecimiento, desarrollo y respuestas al ambiente, por lo tanto, los resultados de investigación son considerados potencialmente aplicables a cualquier otra planta superior. Por otro lado, diferentes técnicas de mutagénesis son aplicables a esta planta, lo cual ha permitido generar numerosas colecciones de líneas mutantes (Azpiroz-Leehan y Feldman, 1997; Bent, 2000). En algunas

de estas colecciones generadas por inserción se han secuenciado las regiones que flanquean al T-DNA o al transposón utilizado. Esto permite conocer fácilmente el efecto fenotípico de la alteración de un gen concreto. Estas colecciones incluyen más de 200.000 líneas (Kuromori *et al.*, 2004; Rosso *et al.*, 2003; Pan *et al.*, 2003).

## I.2 Estructura corporal y ciclo de vida

*Arabidopsis thaliana* (L.) Heynh. es una planta vascular, angiosperma, de la clase de las Dicotiledóneas y que pertenece a la familia de las Brassicáceas o Crucíferas (Strasburger *et al.*, 1994). La familia de las Brassicáceas está formada por unos 200 géneros y 2000 especies, algunas de importancia económica. Entre ellas destacan las plantas forrajeras y olerícolas como el nabo (*Brassica rapa* var. *rapa*), el rábano (*Raphanus sativus*) y las coles (*Brassica oleracea*), plantas oleaginosas como la colza (*Brassica napus*), especies como la mostaza blanca (*Sinapis alba*) y negra (*Brassica nigra*) y plantas ornamentales como el alhelí (*Matthiola incana*) (Strasburger *et al.*, 1994).

*Arabidopsis thaliana* es una pequeña planta herbácea que presenta el ciclo de vida característico de las plantas con flores, con alternancia entre generaciones gametofítica y esporofítica, en el que pueden distinguirse tres etapas: la gametogénesis, la embriogénesis y el desarrollo vegetativo, que se inicia con la germinación y finaliza al aparecer las flores (Jürgens y Mayer, 1994). La duración de cada una de estas etapas es de una, dos y tres semanas, respectivamente, dependiendo además del ecotipo y de las condiciones ambientales (Rédei, 1970).

El desarrollo vegetativo comienza con la germinación. El crecimiento del embrión rompe la cubierta de la semilla, y se produce la emergencia, con la elongación de la radícula y la posterior expansión de los cotiledones. A este estadio inicial se le conoce como estado de plántula (Figura 1.A). Posteriormente, la radícula emergente crece y se diferencia, dando lugar a la raíz. Las estructuras aéreas se desarrollan con la aparición de las hojas verdaderas que inicialmente forman la roseta (Figura 1.B). Las hojas son

simples, de elípticas a ovales, con los bordes enteros y de hasta 2 cm de largo por 0,5 cm de ancho.

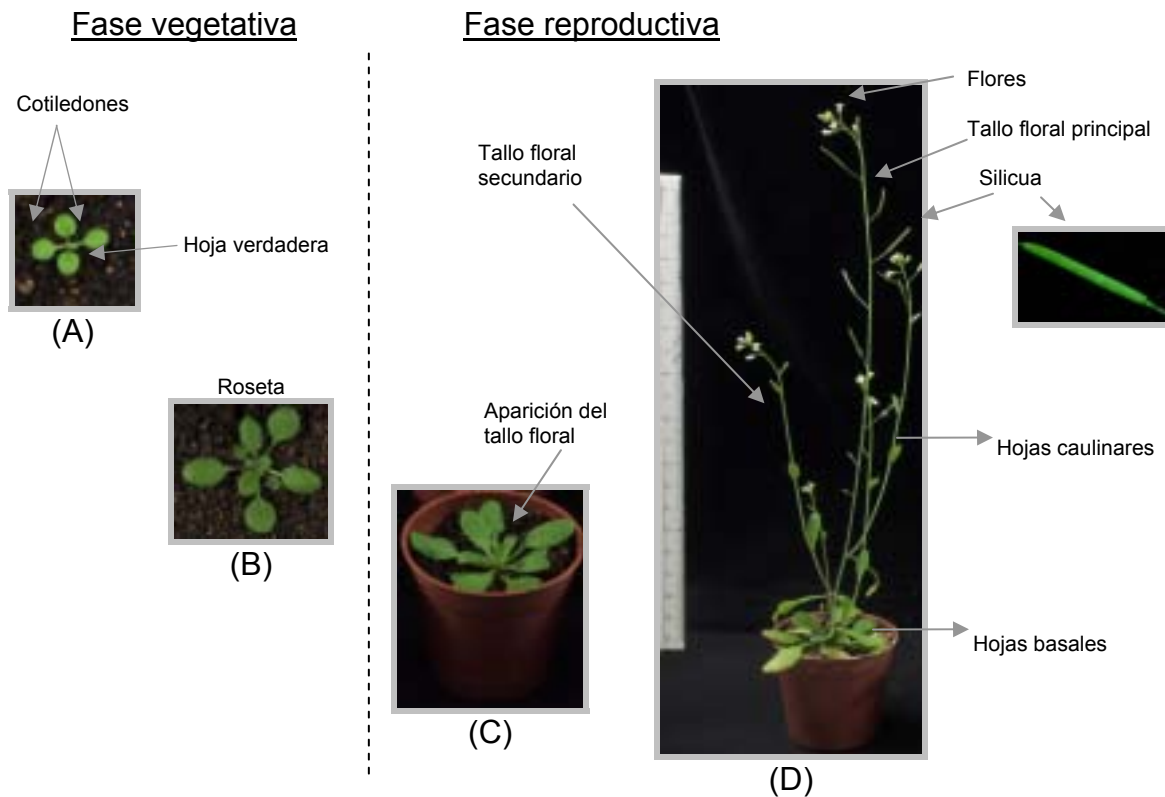


Figura 1.- Fases de desarrollo y órganos de una planta de *Arabidopsis thaliana*, ecotipo Columbia-0

La segunda etapa corresponde a la fase reproductiva, en la que a partir del tallo basal de la roseta, emerge un tallo floral principal con una altura comprendida entre los 10 y 30 cm (Figura 1.C) y algunos tallos florales secundarios. En estos tallos se sitúan las hojas caulinares, más pequeñas que las de la roseta y sésiles (carentes de pecíolo) (Figura 1.D). Cada uno de los tallos desarrolla inflorescencias terminales y secundarias, con flores dispuestas en racimos (Figura 1.D). Las flores se van separando unas de otras a medida que el tallo crece, por lo que los pedúnculos de los frutos maduros estarán separados del orden de un centímetro entre ellos. Las flores son hermafroditas, de unos 0,5 cm de diámetro, normalmente con cuatro pétalos blancos, espatulados, 4 sépalos y 6 estambres. Se distinguen en ella sin dificultad todos los órganos florales.

La generación esporofítica comienza con la formación de un cigoto diploide tras la fusión de los gametos en la fertilización. Una vez fecundadas,



las flores dan lugar a un fruto dehiscente denominado silicua, de unos 3 cm de longitud y 1 mm de anchura, cilíndrico y un poco arqueado. La silicua contiene dos cavidades en las que se alojan las semillas ovoideas en hilera, sin tocarse entre ellas, en número elevado (de 30 a 60 por silicua), y de unos 0,5 mm de longitud. Durante el desarrollo de las semillas se suceden las distintas fases de la embriogénesis, que culminan con la formación del embrión maduro, que presenta las características básicas del plan corporal de la planta, como son la organización radial de los tejidos y el establecimiento de los elementos básicos del patrón apical-basal (los meristemos caulinar y radicular, el hipocótilo y los cotiledones), que serán profundizados en el capítulo I.6.1.

### I.3 El genoma de *Arabidopsis thaliana*

El genoma haploide de *Arabidopsis thaliana* contiene unas 157 Mb (Bennett *et al.*, 2003) divididas en cinco cromosomas (The Arabidopsis Genome Initiative, 2000). Este tamaño es similar al de *Drosophila melanogaster* (180 Mb; Adams *et al.*, 2000) y *Caenorhabditis elegans* (100 Mb; <http://www.wormbase.org/>), y aproximadamente trece veces mayor al de *Saccharomyces cerevisiae* (12 Mb; Goffeau *et al.*, 1996), pero considerablemente menor que el de otras plantas como el arroz (466 Mb; Goff *et al.*, 2002; Yu *et al.*, 2002), maíz (2.500 Mb; Bennetzen, 2002), cebada (4.900 Mb; Bennetzen, 2002) o trigo (16.000 Mb; Adam, 2000). Dentro de la familia de las crucíferas, el genoma de la col (*Brassica oleracea*) alcanza unas 760 Mb, el de la mostaza (*Sinapis alba*) 490 Mb, y el de *Cardamine amara* es de sólo 49 Mb (Hall *et al.*, 2002).

El número de total de genes estimado en el genoma de *Arabidopsis thaliana* es de algo más de 26.000 (Berardini *et al.*, 2004), un valor superior comparado con los genomas de *Drosophila melanogaster* (13.601; Adams *et al.*, 2000; Wigge y Weigel, 2001) y *Caenorhabditis elegans* (19.099; The *Caenorhabditis elegans* Sequencing Consortium, 1998; Wigge y Weigel, 2001) y cercano al de la especie humana, con poco más de 30.000 genes (International Human Genome Sequencing Consortium, 2001; Venter *et al.*, 2001; Pennisi, 2003). Esto supone que el genoma de *Arabidopsis thaliana* posee una densidad de genes de 1 gen cada 6 kb (The Arabidopsis Genome Initiative, 2000). La longitud promedio de los genes es de 2 kb.

La identificación de todos los genes que contiene un genoma no es un problema sencillo (Urbánek *et al.*, 2005). Programas informáticos permiten detectar genes basándose en la conservación de ciertas secuencias, pero dichos genes no pueden ser considerados más que hipotéticos hasta que existan pruebas experimentales de su existencia y estructura. Una de estas pruebas es la existencia de mRNAs (cDNAs). Se han invertido considerables esfuerzos en la secuenciación masiva de moléculas de cDNA, las denominadas ESTs (*Expressed Sequence Tags*). En la base de secuencias dbEST existen 421.027 entradas para esta especie (versión 200106, 20 de enero de 2006) ([http://www.ncbi.nlm.nih.gov/dbEST/dbEST\\_summary.html](http://www.ncbi.nlm.nih.gov/dbEST/dbEST_summary.html)). A pesar de ello, no se han detectado ESTs correspondientes a muchos de los genes predichos *in silico* a partir de la secuencia genómica (Yamada *et al.*, 2003). Se calcula que existen ESTs para solo unos 16.115 genes (Rudd, 2003). El resto de genes continúan siendo hipotéticos. Se puede suponer que un número de genes todavía escapa a la verificación experimental debido a su baja abundancia de transcritos y/o a la severa restricción espacio-temporal del patrón de expresión, pero en otros casos pueden tratarse de pseudogenes o genes no activos, o, simplemente, de errores de predicción.

El genoma de *Arabidopsis* presenta un alto grado de repeticiones de distinto tipo. Por una parte, más de la mitad del genoma está duplicado (Figura 2) ([http://www.tigr.org/tdb/e2k1/ath1/arabidopsis\\_genome\\_duplication.shtml/](http://www.tigr.org/tdb/e2k1/ath1/arabidopsis_genome_duplication.shtml/)). Por otro lado, una elevada proporción de los genes de *Arabidopsis* están duplicados y organizados en el genoma en forma de repeticiones en tándem (Figura 3): un total de 1.528 grupos de genes en tándem que contienen en su conjunto 4.140 genes, con hasta 23 miembros adyacentes (Bevan *et al.*, 2001).

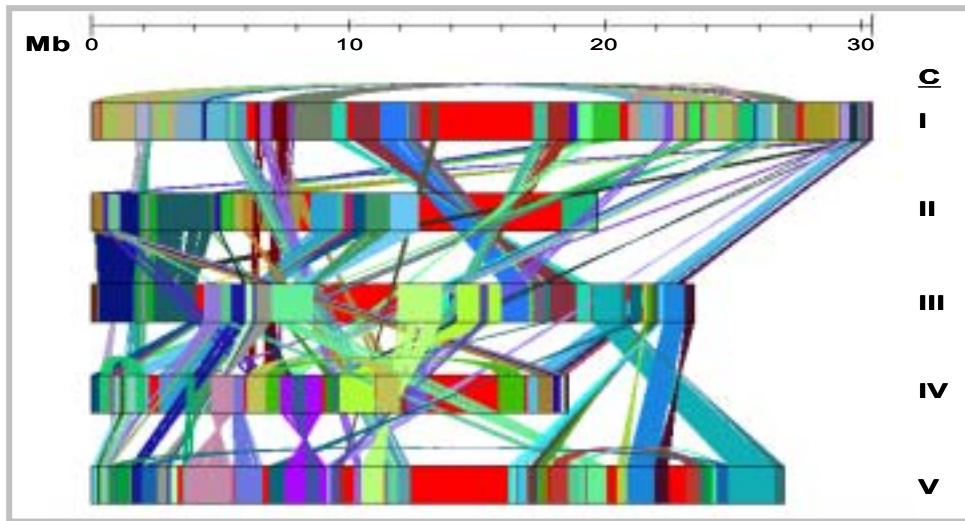


Figura 2.- Análisis del genoma de *Arabidopsis thaliana*. Regiones duplicadas en el genoma de *Arabidopsis*. Los cromosomas (I a V) están representados mediante barras horizontales. Los segmentos sin duplicar son de color rojo. Las bandas de cada color conectan los segmentos duplicados. Los segmentos duplicados en orientación reversa están conectados con bandas de color giradas. Escala del tamaño de los cromosomas expresados en Mb. Adaptado de [http://www.tigr.org/tdb/e2k1/ath1/arabidopsis\\_genome\\_duplication.shtml/](http://www.tigr.org/tdb/e2k1/ath1/arabidopsis_genome_duplication.shtml/)

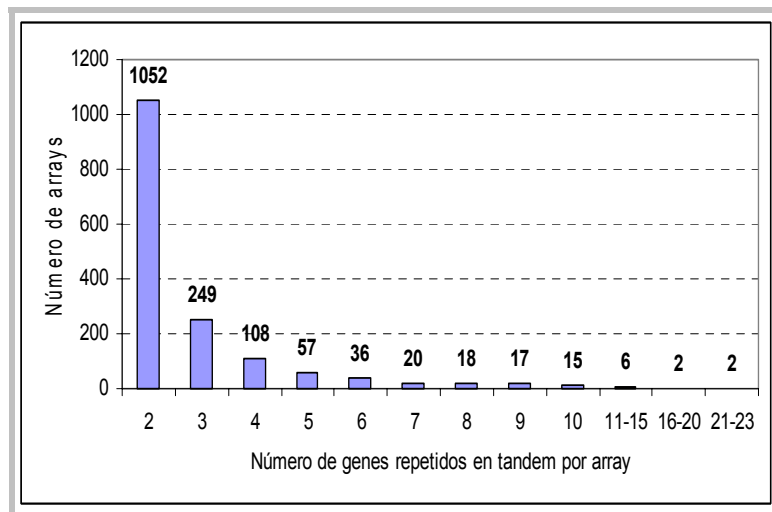


Figura 3.- Distribución de *arrays* de genes repetidos en tándem. Los *arrays* de genes repetidos en tándem, fueron identificados usando el programa BLASTP con un umbral de  $E < 10^{-20}$ . Se toleró un gen sin relación entre los miembros de cada grupo. El histograma dio el número de grupos en el genoma que contiene de 2 a  $n$  unidades de genes similares en tándem. Adaptado de The Arabidopsis Genome Initiative (2000), con modificaciones

Entre las clases funcionales de los genes identificados en el genoma de *Arabidopsis thaliana* destacan los de metabolismo, transcripción, crecimiento, división celular, entre otros, sin embargo, existe una alta proporción de genes cuya función es desconocida (36 %) (Figura 4) (Berardini *et al.*, 2004).

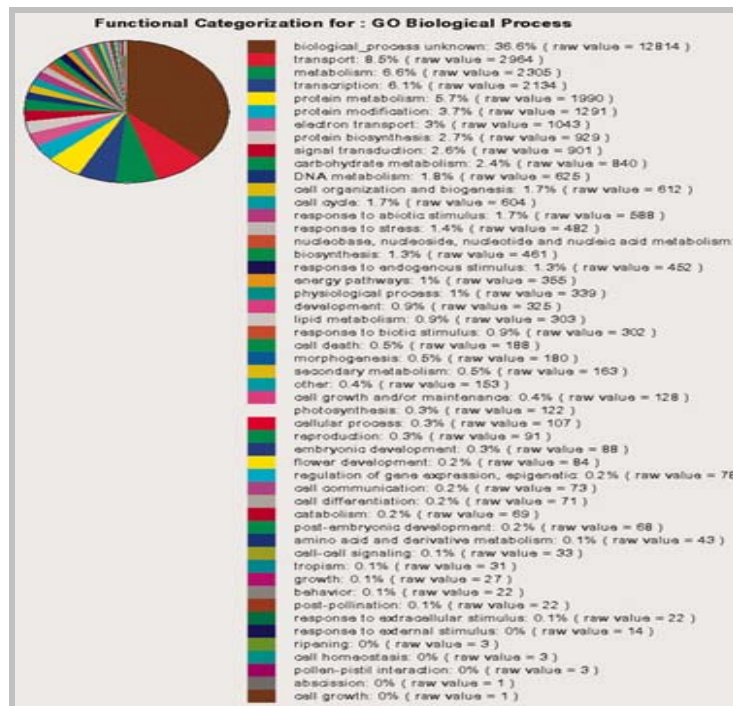


Figura 4.- Análisis del genoma de *Arabidopsis thaliana*. Proporción de genes predichos en diferentes categorías funcionales. Adaptado de Berardini *et al.*, 2004.

El genoma de *Arabidopsis* contiene unas 7.500 familias de genes, considerando como familia de genes a aquél conjunto de genes que codifican proteínas cuyas secuencias son similares (The *Arabidopsis* Genome Initiative, 2000). El número de tipos de proteínas distintas que se codifican por los genes de *Arabidopsis* se estima en unas 11.600, similar a otros organismos multicelulares. Por tanto, un 65% de los tipos de proteínas son codificados por familias de genes.

#### 1.4 Herramientas informáticas para *Arabidopsis thaliana*

Existe un amplio repertorio de recursos informáticos disponibles para el estudio de la biología de *Arabidopsis thaliana* accesibles vía Internet. Entre los de mayor relevancia podemos señalar la base de datos TAIR (The *Arabidopsis* Information Resource), que contiene abundante información sobre genes, familias de genes, marcadores, polimorfismos, mapas genéticos y físicos, secuencias de DNA y rutas metabólicas, así como las publicaciones y los grupos de investigación relacionados con esta especie (<http://www.arabidopsis.org/>); también, las bases de datos que reúnen la información y aportan las semillas disponibles de las líneas mutantes asociadas

a un determinado gen como por ejemplo, NASC (The Nottingham Arabidopsis Stock Centre) y ABRC (Arabidopsis Biological Resource Center) (<http://arabidopsis.info/>) y la herramienta SIGnAL (The Salk Institute Genome Analysis Laboratory) (<http://signal.salk.edu/cgi-bin/tdnaexpress>) o bien, la base de datos de inserciones (<http://atidb.org/>); la información acerca de las secuencias que han sido depositadas a partir de las diferentes genotecas de *Expressed Sequences Tags* (ESTs) generadas en esta especie son reunidas en la base de datos de TIGR (The Institute for Genomic Research) (<http://www.tigr.org/tdb/e2k1/ath1/ath1.shtml>) junto a la de otras especies y organismos.

Existe una amplia base de datos de resultados de *microarrays* disponibles para la mayoría de los genes de Arabidopsis. Uno de ellos es *Gene Expression Visualization de AtGenExpress* (<http://www.weigelworld.org/resources/microarray/AtGenExpress>). Mediante este programa se puede visualizar la intensidad de expresión de los genes de *Arabidopsis thaliana* en diferentes órganos y estados fenológicos de la planta (Schmid *et al.*, 2005). Por otra parte, la base de datos Genevestigator<sup>®</sup> (<https://www.genevestigator.ethz.ch/>) que también permite analizar la intensidad de expresión de los genes bajo diferentes condiciones de estrés y en diferentes órganos, basándose en la información disponible en 2.317 matrices (Zimmermann *et al.*, 2004, 2005).

### 1.5 Desarrollo del polen en *Arabidopsis thaliana* (microsporogénesis)

En las plantas con flor el desarrollo reproductivo masculino requiere la formación del estambre y la diferenciación de los tejidos de la antera. En la antera, la meiosis produce las micrósporas, que se convierten más tarde en los granos del polen (Figura 5.A) (Ma, 2005). El citoplasma de la célula de polen hay abundantes orgánulos y reservas: retículo endoplasmático, dictiosomas, plástidos con almidón que se consume en la formación del tubo polínico, lípidos, proteínas y vitaminas.

El grano de polen se dispersa cuando las paredes de las anteras se abren, permitiendo que la polinización ocurra. Una vez liberados, los granos de

polen están expuestos a condiciones extremas y a menudo durante largo tiempo. La protección de su contenido está asegurada por la presencia de una pared muy resistente llamada esporodermis. De afuera hacia dentro se pueden distinguir dos capas: exina e intina. La intina envuelve al protoplasma, es delicada, poco resistente y constituida de celulosa y pectina. La exina está constituida por esporopolenina, un polímero de carotenos y sus ésteres, sustancia químicamente muy resistente y solo degradable por oxidación (Gavarayeba, 1996). La exina muestra un alto grado de diferenciación estructural en las Angiospermas (Figura 5.B y C). En las especies más primitivas puede ser amorfa pero en las demás se pueden distinguir 2 partes: la nexina, interna, homogénea, y la sexina, externa, que es la porción esculturada. Consta de bastones o báculas que pueden unirse entre sí por los extremos formando el tectum (Fahn, 1982).

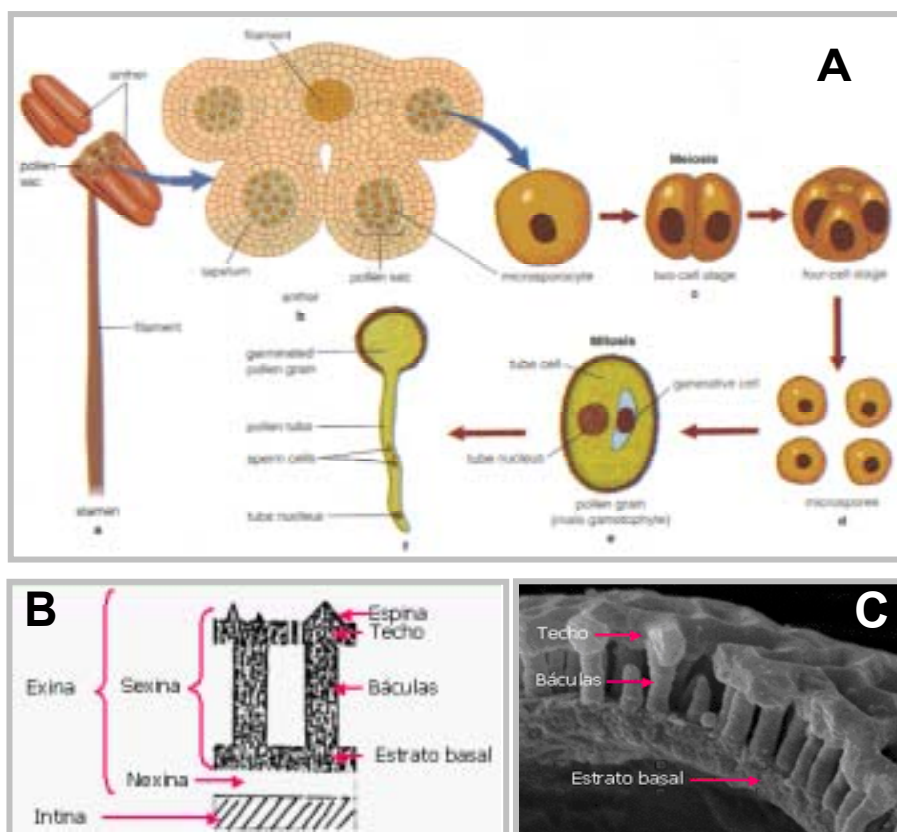


Figura 5.- Desarrollo del polen. A. Representación esquemática de las distintas fases del desarrollo del polen. B. Representación esquemática de un corte transversal de la esporodermis del grano de polen. C. Aspecto de un corte transversal de la esporodermis del grano de polen observado mediante microscopía electrónica.

Estudios moleculares han identificado una serie de genes que se expresan durante el desarrollo del polen (Ma, 2005). La secuenciación de ESTs (Engel *et al.*, 2003) y los análisis proteómicos (Mayfield *et al.*, 2001; Imin *et al.*, 2001) han proporcionado gran cantidad de información. Los análisis genéticos han demostrado que la función de algunos de estos genes es especificar la identidad del polen, la división de las células de la antera y la meiosis, la regulación de la diferenciación del polen o la inducción de la dehiscencia de la antera. Estos genes codifican una variedad de proteínas, incluyendo reguladores de transcripción, proteínas de transducción de señal, reguladores de la degradación de proteínas, enzimas para la biosíntesis de hormonas, proteínas relacionadas con la muerte celular programada de la antera o histonas de expresión específica en el polen (Ueda y Tanaka, 1995). La identificación de genes responsables de líneas mutantes en el desarrollo del polen también ha proporcionado importante información sobre este proceso (Ross y Murphy, 1996; Ross *et al.*, 1997; Bhatt *et al.* 2001; Wilson y Yang, 2004; Rogers, 2006).

### I.6 Desarrollo embrionario en *Arabidopsis thaliana*

Se entiende por embriogénesis al proceso mediante el cual se forma un embrión capaz de convertirse en una planta adulta a partir de una única célula, el cigoto (Figura 6) (Harada, 1999).

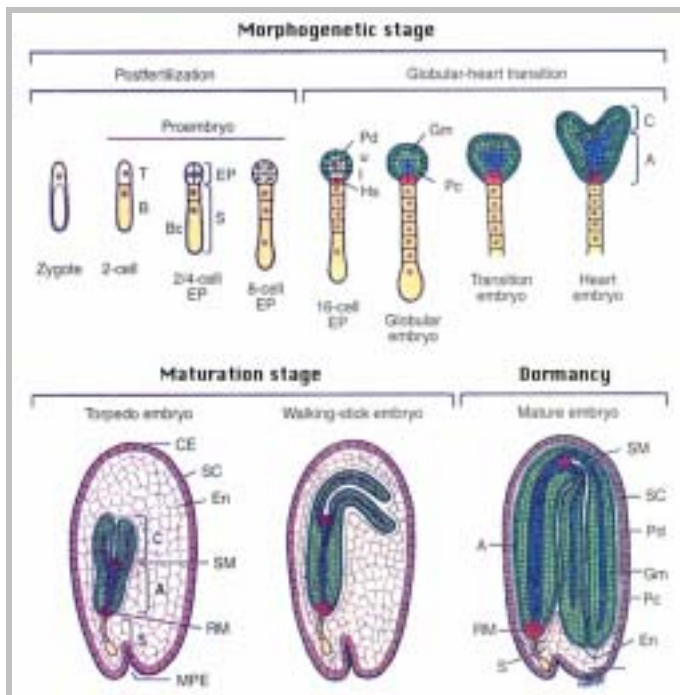


Figura 6.- Una visión generalizada de la embriogénesis vegetal. Representaciones esquemáticas de las etapas de la embriogénesis basada en estudios de desarrollo del embrión de *Arabidopsis* mediante microscopía óptica. Abreviaturas: T, célula terminal; B, célula basal; EP, embrión; S, suspensor; Bc, célula basal del suspensor; Pd, protoderma; u, región superior; l, región inferior; Hs, hipófisis; Pc, procambium; Gm, meristemo base; C, cotiledón; A, eje; MPE, extremo micropilar; CE, extremo chalazal; SC, cubierta de la semilla; En, endospermo; SM, meristemo apical; y RM, meristemo radical. Adaptado de Goldberg *et al.*, (1994) con modificaciones.



### I.6.1 Formación del patrón: etapa morfogénica

Los pasos iniciales de la embriogénesis se caracterizan por la formación de los diferentes órganos y tejidos que constituirán el embrión (etapa morfogénica). Uno de los principales procesos dentro de la formación inicial del embrión es el establecimiento de los patrones de simetría, es decir, de la forma básica del embrión. Esto incluye el patrón radial (interno-externo) y del patrón apical-basal (parte aérea-raíz) (Jürgens, 2001; Willemsen y Scheres, 2004).

La primera división del cigoto ya es asimétrica, dando lugar a dos células de diferente tamaño y densidad (Figura 6). La célula apical, más pequeña y densa, dará lugar a la mayor parte del embrión propiamente dicho, mientras que la basal, más grande y vacuolada, dará lugar al suspensor y parte de la raíz (Berleth, 1998). La célula superior del suspensor forma la hipófisis, que da lugar al extremo de la región radicular del eje embrionario (Mansfield y Briarty, 1990; Dolan *et al.*, 1993) y, por otra parte, tiene la misión de proyectar el embrión en el tejido maternal y proveerlo de nutrientes y factores de crecimiento (Yeung y Meinke, 1993; Souter y Lindsey, 2000).

La célula apical densa sufre diferentes divisiones hasta formar el embrión globular, de forma esférica. Aquí tiene lugar el primer signo del establecimiento del patrón radial del embrión. Las divisiones tangenciales que tienen lugar en el estadio de 8 células (octante) separan a las células externas (protoderma) de las internas. El protoderma dará lugar a la epidermis mediante divisiones posteriores de tipo anticlinal, mientras que las células interiores generaran los tejidos internos (Laux y Jürgens, 1997).

La diferenciación de los órganos y tipos de tejidos embrionarios comienza durante el período de transición de globular a corazón (Figura 7). En este momento cambia la morfología del embrión, estableciéndose la polaridad apical-basal (Sheridan y Clark, 1993; Jürgens *et al.*, 1994; Hemerly *et al.*, 2000; Elster *et al.*, 2000). La región apical (Figura 7 *ad* (verde)) da origen al meristemo apical y cotiledones, y la región central (Figura 7 *cd* (amarillo)) contribuye a la formación del cotiledón, el hipocótilo, la radícula y el meristemo radical.



Finalmente, a partir de la región basal se desarrolla el centro quiescente y el caliptrógeno (Figura 7 *hy* (azul); Figura 8) (West y Harada, 1993; Souter y Lindsey, 2000; Jürgens, 1995; Laux y Jürgens, 1997; Hudson, 2000).

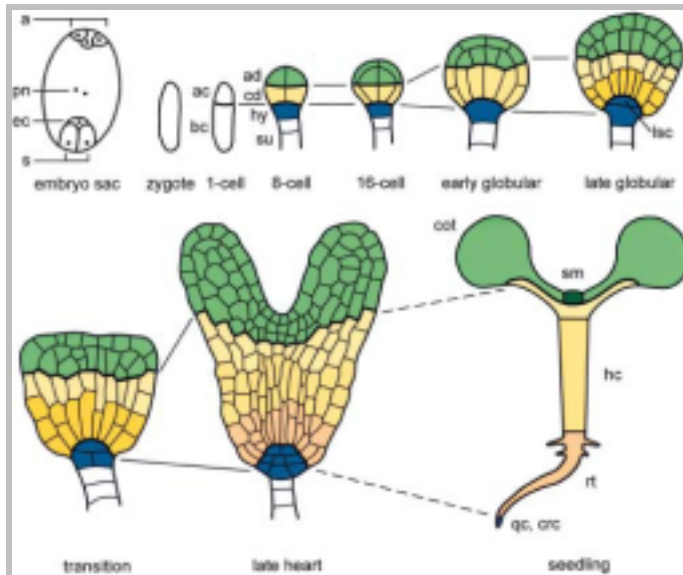


Figura 7.- Desarrollo Apical-Basal del embrión de *Arabidopsis*. Esquema de secciones medias longitudinales. Las líneas gruesas superiores e inferiores representan los límites entre los descendientes de las células hijas apical y basal y entre los dominios embrionarios apical y central, respectivamente. Abreviaturas: a, antípodas; ac, célula hija apical; ad, dominio embrionario apical; bc, célula hija basal; cd, dominio embrionario central; cot, cotiledones; crc, caliptrógeno; ec, óvulo; hc, hipocótilo; hy, hipófisis; lsc, célula lenticular; pn, núcleo polar; qc, centro quiescente; rt, raíz; sm, meristemo apical; su, suspensor. Adaptado de Laux *et al.* (2004), con modificaciones.

El establecimiento del patrón apical-basal y del radial corresponde al aspecto organizativo de la embriogénesis, es decir, los diferentes órganos, tejidos y tipos celulares deben formar un contexto coherente a nivel estructural y funcional (Jürgens *et al.*, 1994). Una vez establecida la organización del cuerpo, mediante procesos de morfogénesis se adquiere la forma final del embrión en general y de los diferentes órganos en particular. Debido a que la célula vegetal posee una pared celular que no les permite migrar, el crecimiento y la diferenciación son fundamentales mecanismos en la orientación y número de las divisiones celulares y en la diferente expansión celular (Jürgens *et al.*, 1991; Jürgens, 1992; Hemerly *et al.*, 1995, 1999; Sheridan, 1995).

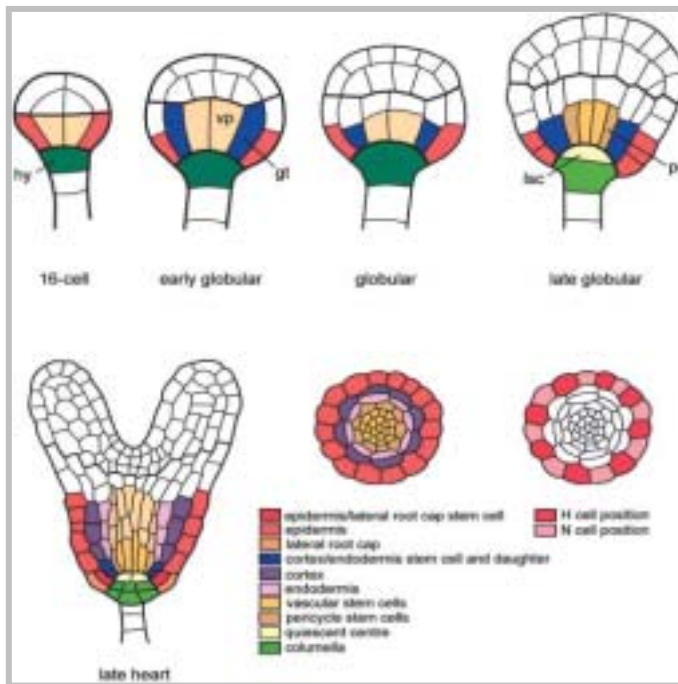


Figura 8.- Desarrollo del patrón radial del embrión de Arabidopsis. La fila superior y la ilustración abajo a la izquierda muestran esquemas de secciones longitudinales. Las otras ilustraciones en la fila inferior muestran esquemas de sección de corte a través de la raíz. Las líneas gruesas representan los límites entre los descendientes de las células hijas apical y basal y entre los dominios embrionarios, respectivamente. Los tipos celulares son mostrados en color como se indica. Las células vasculares y del periciclo son mostradas en colores más claros que las células del tallo. Abreviaturas: gt, tejido basal; hy, hipófisis; lsc, célula lenticular; pc, periciclo; vp, primordio vascular. Adaptado de Laux *et al.* (2004), con modificaciones.

### 1.6.2 Maduración y latencia

Una vez establecida la organización del cuerpo y las capas celulares del embrión comienza el estado de maduración que está caracterizado por la expansión de órganos, la acumulación de proteínas de reserva y la preparación de la semilla para la latencia (Goldberg *et al.*, 1994). La mayor parte de los productos de reserva son acumulados en los cotiledones. Hacia el fin del estado de maduración el embrión ha alcanzado su tamaño máximo, las capas que rodean a las semillas son deshidratadas, las actividades metabólicas disminuyen y comienza el período de latencia. Previamente a la desecación se produce la acumulación de una serie de proteínas cuya función es proteger al embrión durante este proceso (Delseny *et al.*, 2001).

### 1.6.3 Control genético de la embriogénesis

El desarrollo del embrión está bajo un estricto control genético regulado por una gran cantidad de genes que han de funcionar de manera coordinada y organizada, tanto temporal como espacialmente (van Lijsebettens y van Montagu, 2005). La diferenciación celular y de tejidos es dependiente de patrones específicos de expresión de genes. Las transiciones de desarrollo están acompañadas por cambios en la expresión de genes (Doebley y Lukens, 1998; Gatehouse *et al.*, 1986; White *et al.*, 2000)

El estudio de mutantes es una estrategia general para identificar genes esenciales para el correcto desarrollo del embrión y la semilla (Goldberg *et al.*, 1994; Meinke, 1995). También, permite determinar que procesos funcionan de manera independiente, cuales están relacionados entre sí y las posibles jerarquías en los diferentes procesos (Meinke, 1995). Entre los distintos tipos de agentes utilizados para la obtención de mutantes en *Arabidopsis thaliana*, y en plantas en general, destaca el metanosulfonato de etilo (EMS) que es un agente químico alquilante. Entre los mutágenos físicos frecuentemente empleados cabe destacar a los rayos X y los neutrones rápidos, que suelen inducir deleciones y otras aberraciones cromosómicas (Shirley *et al.*, 1992; Bruggemann *et al.*, 1996). La inserción de elementos transponibles es otra herramienta que se ha utilizado para generar mutaciones. Aunque *Arabidopsis thaliana* cuenta con transposones endógenos (Peleman *et al.*, 1991; Tsay *et al.*, 1993; Miura *et al.*, 2004), han demostrado ser de mayor utilidad los sistemas que emplean transposones heterólogos procedentes del maíz, el Ac/Ds (*Activator/Dissociation*; Bancroft *et al.*, 1992) y el En/I (*Enhancer/Inhibitor*), también conocido como Spm-dSpm (*Supressor-mutator*; Aarts *et al.*, 1995). Sin embargo, el mutágeno insercional por excelencia es el T-DNA (*transfer DNA*), un segmento del plásmido Ti (*tumor inducer*, inductor de tumores) que se integra en el genoma nuclear de las plantas infectadas por la bacteria *Agrobacterium tumefaciens* (Parinov y Sundaresan, 2000; Bouche y Bouchez, 2001; Thorneycroft *et al.*, 2001; Ostergaard y Yanofsky, 2004).

El estudio de los mutantes de *Arabidopsis* ha evidenciado que las regiones apical, central y basal definen grupos celulares que expresan diferentes genes. Entre estos mutantes se pueden citar *gurke* (GK) (Torres Ruiz *et al.*, 1996), *fackel* (FK) (Mayer *et al.*, 1991), *monopteros* (MP) (Berleth y Jürgens, 1993), *gnom* (GN) (Mayer *et al.*, 1993; Meinke, 1995) y *bodenlos* (BDL) (Hamann *et al.*, 1999), que presentan alteración exclusiva en una o dos regiones del patrón apical-basal. De este modo, los mutantes *gurke* no forman cotiledones y fallan en la formación del meristemo apical, mientras que en los mutantes *fackel*, el tejido del hipocótilo no separa la región apical de la región basal, es decir, presenta deleciones en la parte central del eje generando plántulas con los cotiledones unidos directamente a la raíz (Barceló *et al.*,

2003). Por su parte, los mutantes *gnom* carecen de dominios apical y basal, por lo que presentan plántulas sin raíces ni cotiledones perdiendo en casos extremos la polaridad axial por completo (Figura 9). Una caracterización más detallada de estos mutantes indica que la mutación afecta primero a la región correspondiente, pero después se puede observar el efecto en las células vecinas (Hudson, 2000; Souter y Lindsey, 2000). Estos resultados sugieren que los elementos del patrón apical-basal son establecidos por interacciones celulares dependiendo de la posición y que esta información de posición, permite a las células activar una expresión de genes coordinada y diferenciada (Jürgens, 1995; Mayer y Jürgens, 1998; Hudson, 2000; Souter y Lindsey, 2000; Jürgens, 2001).

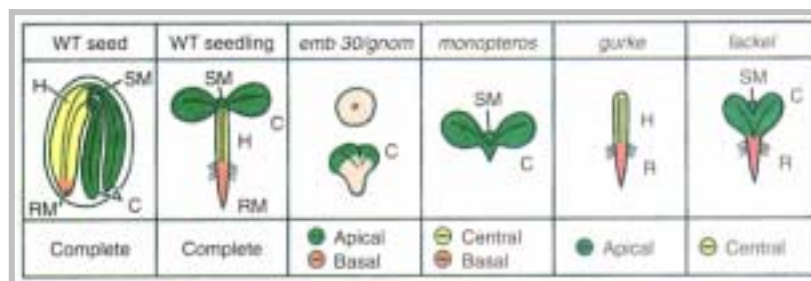


Figura 9.- Representación esquemática de mutantes de Arabidopsis. Los colores verde, amarillo y naranja señalan las regiones apical, central y basal, respectivamente. Se indican las regiones que faltan. Abreviaturas: WT, silvestre; RM, meristemo radical; SM, meristemo apical; C, cotiledón; h, hipocotilo; R, raíz. Adaptado de Goldberg *et al.* (1994).

Se han identificado genes que tienen alteraciones en el patrón radial del embrión como ocurre con el gen *short root*, cuyos mutantes presentan plántulas sin endodermis, o el gen *scarecrow*, en cuyos mutantes no hay separación entre endodermos y tejido cortical (Sabatini *et al.*, 2003; Paquette y Benfey, 2005). Otro mutante interesante es el mutante *knolle* (KN), en el que las células internas no están separadas del protodermo dentro del pro-embrión y de este modo, la plántula del mutante carece de la capa de epidermis característica (Mayer *et al.*, 1991). Además, el mutante *knolle* tiene paredes celulares incompletas y orientadas de forma incorrecta, lo que produce la continuidad entre las células interiores y el hipotético protodermo (Laux y Jürgens, 1997; Hudson, 2000). El resultado del análisis del mutante *knolle* sugiere que una separación física de las células es necesaria para la propia diferenciación celular, quizás para proveer una comunicación regulada de célula a célula.

Por otra parte, se han identificado y caracterizado otros mutantes en los que el desarrollo del embrión se detiene y el suspensor prolifera de manera continuada, como el mutante *raspberry* (Yadegari *et al.*, 1994), el mutante *sus* (Schwartz *et al.*, 1994), o el *twn2* (Vernon y Meinke, 1994; Zhang y Sommerville, 1997). Las células de los suspensores de estos mutantes recuerdan las del embrión, ya que acumulan productos de reserva y adquieren características únicas del embrión. También se han detectado mutantes con óvulo alterado (*bell1*, *fis1*, *cuc1*), endospermo alterado (*fie*, *mea*, *dme*), o desarrollo tardío del embrión alterado (*abi3*, *lec1*, *fus3*, *twn1*) (West *et al.*, 1994; Reiser *et al.*, 1995; Hemerly *et al.*, 2000; Vernon *et al.*, 2001; Choi *et al.*, 2002).

El análisis de mutantes ha proporcionado valiosa información sobre la manera en que se regula el desarrollo embrionario (McElver *et al.*, 2001). Se calcula que en *Arabidopsis* la mutación de unos 250 genes produce letalidad a nivel del embrión (Tzafrir *et al.*, 2004). Sin embargo, no todos los genes con un papel importante durante la embriogénesis se pueden identificar mediante el análisis de mutantes, como es el caso de los genes con una función redundante o cuya función es requerida durante la formación de los gametos y que, por lo tanto, no podrán generar embriones (Meinke 1991, 1995). Por el contrario, no todos los genes cuya mutación bloquea el desarrollo del embrión son genes relacionados directamente con la embriogénesis, ya que genes necesarios para el funcionamiento basal de las células pueden provocar también letalidad a nivel del embrión.

## Capítulo II. Dominios proteicos

Una forma de predecir la función de un gen es determinar la presencia y organización de dominios proteicos conservados (Sessions *et al.*, 2002). Se entiende por dominio proteico una porción de una proteína con estructura terciaria definida y, en general, asociado a alguna función. El creciente número de secuencias de proteínas depositadas en las bases de datos ha revelado la existencia de motivos de secuencia comunes algunos de los cuales han sido asociados con funciones determinadas. Se han identificado 680 dominios en la base de SMART (Simple Modular Architecture Research Tool; <http://www.smart.embl-heidelberg.de/>). Esto permite la clasificación de las proteínas en base a la presencia de motivos (Sedgwick y Smerdon, 1999).

### II.1 Proteínas con dominios transmembrana

Los dominios transmembrana están compuestos por entre 15 y 30 residuos mayoritariamente hidrofóbicos y se encuentran en proteínas que se unen a alguna membrana celular. Aproximadamente 4.589 genes de *Arabidopsis thaliana* (18%) contienen dos o más dominios transmembrana (Ward, 2001; ARAMEMNON: <http://botanik.uni-koeln.de/>, Schwacke *et al.*, 2003). Las proteínas integrales de membrana pueden tener importantes funciones en transporte de solutos a través de la membrana, percepción y transducción de señales y actividades biosintéticas y metabólicas localizadas en membrana. Una proteína puede tener múltiples dominios transmembrana (ej. 20 en TM20 de maíz) (Stiefel *et al.*, 1999). TM20 de maíz es una proteína transmembrana cuya posible función es el transporte polar de hormonas (Stiefel *et al.*, 1999; Jahrman *et al.*, 2005). Esta proteína se identificó en un mutante tipo *dek* (*defective kernel*) de maíz denominado *lachrima*, cuyo embrión presenta numerosas alteraciones en los estadios tempranos de la embriogénesis. Al comparar la secuencia proteica de TM20 con las bases de datos de otras especies vegetales se encontró una proteína de organización muy similar en arroz (*Oriza sativa* L.). En *Arabidopsis thaliana*, no existe ningún gen que codifique una proteína con 20 dominios transmembrana pero si uno

que codifica una proteína con cierta similitud pero con sólo cuatro dominios transmembrana (Jahrmann, 2002).

## II.2 Repeticiones Anquirina

En 1987 Breeden y Nasmyth describieron la presencia de unas secuencias de 33 aminoácidos repetidas en tándem en dos proteínas reguladoras de ciclo celular de *Schizosaccharomyces pombe* y *Saccharomyces cerevisiae* (CDC10 y SW16). Repeticiones semejantes también estaban presentes en otras proteínas reguladoras como Notch de *Drosophila melanogaster* y LIN-12 de *Caenorhabditis elegans*. Posteriormente, se descubrieron 24 copias de esta repetición en una proteína humana denominada anquirina, relacionada con el citoesqueleto, que fue la que dio el nombre a la repetición anquirina (ANK) (Sedgwick y Smerdon, 1999).

Actualmente se han identificado repeticiones ANK en numerosas proteínas involucradas en muy diversas funciones y pertenecientes a procariotas, eucariotas e incluso virus (Bork, 1993; Rubstov y Lopina, 2000; Lux *et al.*, 1990; Massung *et al.*, 1992). Entre las variadas funciones de las proteínas con repeticiones ANK se puede mencionar la regulación del ciclo celular, enzimas mitocondriales, interacciones de citoesqueleto, traducción de señal o toxinas (Sedgwick y Smerdon, 1999). Por ejemplo, contienen motivos ANK subunidades de factores de transcripción o reguladores de sistemas transcripcionales (cdc10, SW16, SW14, GAPB $\beta$ , NF- $\kappa$ B/p105, I $\kappa$ B $\alpha$ , bcl-3 o CAMTAs), proteínas intrínsecas de membrana que regulan la diferenciación de tejidos (Lin-12, Glp-1, Notch), un determinante sexual de nematodos (Fem-1), proteínas reguladoras de fosfolipasas (Phlb e iPLA<sub>2</sub>), toxinas de arácnidos, intercambiadores iónicos Cl/HCO<sub>3</sub>, Na/Ca, la ATPasa Na/K, el receptor IP<sub>3</sub>, rianodina, canales de Na voltaje-dependientes o proteínas de adhesión celular (familia L1 CAM) (Bennett y Baines, 2001; Givskov *et al.*, 1988; Larsson *et al.*, 1998; Kiyatkin *et al.*, 1993; Michaely *et al.*, 2002, Chang y Low, 2003).

El hecho que este motivo este conservado en organismos tan distantes evolutivamente como mamíferos y levaduras indica que cumplen un papel importante para la función de las proteínas (Breeden y Nasmyth, 1987). Al

mismo tiempo están presentes en proteínas de muy diferente función lo cual implica que no deben de cumplir una tarea muy específica sino más bien de carácter general (Bork, 1993; Sedgwick y Smerdon, 1999; Klimyuk *et al.*, 1999), mientras que la presencia de otros dominios de señal, interacción proteína-proteína y catalíticos en las mismas proteínas serían los responsables de la función específica (Mosavi *et al.*, 2002). Existe un número creciente de ejemplos en los que las repeticiones ANK funcionan como lugares de unión proteína-proteína, por lo que se considera que esa es su función principal (Sedgwick y Smerdon, 1999). Esta función ha sido demostrada experimentalmente en unión de proteínas heterólogas y mediante homodimerización (Bork, 1993; Lin *et al.*, 1999). Por ejemplo, LaMarco y colaboradores (1991) mostraron que las repeticiones ANK están involucradas en uniones de subunidades de las proteínas  $\beta$  de unión GABA (GABP $\beta$ ), la proteína I- $\kappa$ B $\alpha$  está casi enteramente compuesta de repeticiones anquirina y es capaz de unir la subunidad de 65 kDa de NF- $\kappa$ B (Haskill *et al.*, 1991), la  $\alpha$ -latrotoxina de la araña viuda negra se asocia a receptores extracelulares a través de las 19 repeticiones ANK (Sudhof, 2001), y las proteínas Su(H) y Deltex se unen mediante repeticiones ANK a Notch (Le Gall y Giniger, 2004).

Estudios en proteínas animales demuestran que la estructura primaria de las repeticiones ANK tiene en promedio 33 aminoácidos (Figura 10) (Bork, 1993). Solo cinco de los 33 residuos tienen un grado alto de conservación y en 13 más se conserva el tipo de aminoácido. Aquellos aminoácidos que permanecen invariables corresponden a posiciones hidrofóbicas los cuales son necesarios para mantener la estructura secundaria (Rohde y Bork, 1993; Bork, 1993; Mosavi *et al.*, 2002).

- t - o t L H h A h - - t t - t h h t - L L t - t - t - - - - -

Figura 10.- Secuencia consenso de la repetición ANK según la determinación realizada por Bork (1993), basado en proteínas animales. -, representa un aminoácido cualquiera; *t* representa un aminoácido polar o que tiende a formar giros; *o* representa serina o treonina; *h* representa un aminoácido hidrofóbico y las letras mayúsculas corresponden a aminoácidos conservados de acuerdo al código de una letra.



Esta repetición ANK básica de 33 aminoácidos se haya repetida en tándem normalmente de 4 a 6 veces, aunque se han encontrado proteínas de entre dos y 29 repeticiones (Kohl *et al.*, 2003). Cada repetición puede estar separada de la siguiente entre cero y 20 aminoácidos (Bork, 1993).

La determinación de la estructura tridimensional de algunas proteínas con motivos ANK ha revelado que dichas repeticiones se pliegan en una serie de hélices  $\alpha$  conectadas por giros de 90 grados (Figura 11). La estructura esta estabilizada por hojas  $\beta$  plegadas antiparalelas formadas entre las repeticiones y por uniones hidrofóbicas en la repetición y entre las repeticiones vecinas. Las hojas  $\beta$  plegadas se proyectan desde los pares de hélices casi en ángulo recto. La capacidad de las repeticiones ANK para unir dianas de proteínas implica contacto entre las puntas de los  $\beta$ -*hairpins*, los que están expuestos hacia el solvente, y la superficie de las hélices encaradas hacia el interior del surco. Los residuos localizados en los extremos de los giros, las zonas más expuestas, corresponden a su vez con la zona terminal de la repetición ANK, que es la de secuencia menos conservada. Se cree que esta zona es la que determina la especificidad de unión de las proteínas al dominio (Bennett y Baines, 2001; Mosavi *et al.*, 2002). Por lo tanto, el plegamiento de las repeticiones ANK juega un importante papel para los variados tipos de funciones que realizan (Mosavi *et al.*, 2002). Estudios previos han demostrado que una sola repetición ANK no puede adoptar una estructura de plegamiento, si no comparte una interfase con otra repetición, por lo tanto, la unidad mínima para un correcto funcionamiento es de dos repeticiones (Zhang y Peng, 2000).



Figura 11.- Estructura tridimensional de un conjunto de repeticiones ANK. Estructura tridimensional del conjunto de repeticiones ANK de la proteína miotrofina de *Rattus norvegicus* (Yang *et al.*, 1998). Los cilindros representan hélices alfa.

### II.3 Proteínas de Plantas con Repeticiones Anquirina

Hasta el momento se han caracterizado muy pocas proteínas de plantas que contengan motivos ANK y de la mayoría de ellas no se conoce la función a nivel molecular. Entre ellas se pueden citar:

- ANK1 de tabaco (*Ankyrin repeat protein 1*) y su homóloga AKR2 de *Arabidopsis* (Yan *et al.*, 2002; Kuhlmann *et al.*, 2003), involucradas en la defensa frente a patógenos.
- CAMTAs (*Calmodulin-binding transcription activators*) en *Arabidopsis* y *Brassica napus*, probablemente implicadas en cascadas de regulación de la transcripción (Bouché *et al.*, 2002).
- Familia AKT de *Arabidopsis* (Sentenac *et al.*, 1992; Ketchum y Slayman, 1996; Pilot *et al.*, 2003) y SKT1, un homólogo en patata (Zimmermann *et al.*, 1998), que codifican canales de potasio dependientes de voltaje de la familia Shaker.
- APKs (*ankyrin protein kinases*), son quinasas descritas en *Medicago* (Chinchilla *et al.*, 2003).
- ART2 de *Arabidopsis*, interviene en la defensa frente a patógenos (Peck *et al.*, 2001).
- CAO, (*chlorophyll a/b binding protein harvesting-organelle specific protein*) descritas en *Arabidopsis* (Klimyuk *et al.*, 1999; Jonas-Straube *et al.*, 2001).
- AKR (*Ankyrin Repeat gene*) de *Arabidopsis*, interviene en los procesos de diferenciación celular asociada a luz (Zhang *et al.*, 1992).
- NPR1 (*non-expresser of PR genes*) de *Arabidopsis*, interviene en el control de la respuesta frente a patógenos SAR (Systemic Acquired Resistance) (Cao *et al.*, 1997, 1998).
- ACBP2 (*cytosolic acyl-CoA-binding*) de *Arabidopsis*, es una proteína de unión a acil-CoA (Chye *et al.*, 2000).
- EMB506 de *Arabidopsis*, es necesaria para el correcto desarrollo embrionario (Albert *et al.*, 1999).
- ACD6, una proteína con repeticiones ANK y dominio transmembrana, es un posible regulador y efector de la señal del ácido salicílico en la respuesta de defensa de *Arabidopsis* (Lu *et al.*, 2003).

Los repeticiones ANK son objetos muy atractivos desde el punto de vista experimental ya sea para evaluar nuestra comprensión sobre la relación secuencia-estructura-estabilidad-función en proteínas, como para el desarrollo de herramientas moleculares para aplicaciones biotecnológicas como, por ejemplo, el reconocimiento molecular específico (Devi *et al.*, 2004).

## **OBJETIVOS**

---

El objetivo general de esta tesis ha sido la identificación y estudio de genes relacionados con el desarrollo temprano del embrión de *Arabidopsis thaliana*. Este trabajo es continuación de trabajos previos del grupo de investigación relacionados con el estudio de la embriogénesis en maíz.

En detalle, los objetivos planteados para esta tesis doctoral se han centrado en dos puntos principales:

- A. Identificación de genes implicados en el desarrollo de la semilla de *Arabidopsis thaliana*
  - A.1.- Caracterización del transcriptoma de semilla inmadura. Obtención de una genoteca de cDNAs a partir de RNA extraído de semillas inmaduras en etapas muy iniciales del desarrollo. Secuenciación de ESTs y evaluación de las características de los genes identificados.
  - A.2.- Identificación de genes que se transcriben específicamente durante el desarrollo inicial de la semilla. Utilización de un sistema de selección *in silico* basado en la presencia de ESTs y en las bases de datos de hibridaciones de micromatrices. Evaluar los resultados a nivel experimental y analizar las características de los genes seleccionados.
  
- B. Genes que codifican proteínas con repeticiones anquirina y dominios transmembrana en *Arabidopsis thaliana* (genes *AtAnkTm*).
  - B.1.- Identificación y clasificación de las proteínas codificadas en el genoma que contengan repeticiones anquirina. Determinación de la secuencia consenso de la repetición anquirina en *Arabidopsis*.
  - B.2.- Caracterización de los genes *AtAnkTm* incluyendo su clasificación y análisis filogenético, determinación de su distribución cromosómica, presencia en otras especies, determinación de los patrones de expresión y análisis de líneas mutantes.
  - B.3.- Caracterización de genes *AtAnkTm* específicos de etapas tempranas del desarrollo embrionario.

## **RESULTADOS**

---

**Capítulo I. Identificación de genes que se expresan específicamente durante el desarrollo temprano de la semilla de *Arabidopsis thaliana***

Una de las metas principales de la biología del desarrollo de las plantas es la determinación de los genes implicados en el desarrollo de la semilla y sus funciones. El uso de mutantes ha generado importantes descubrimientos en este área (McElver *et al.*, 2001; Chaudhury *et al.*, 2001; Meinke *et al.*, 2003; Tzafrir *et al.*, 2004; <http://www.seedgenes.org/>). Sin embargo, la mutagénesis tiene algunas deficiencias. Debido probablemente a la redundancia de los genes, muchas de las inserciones no producen ningún fenotipo perceptible. Por otro lado, los genes cuya interrupción produce alteraciones en el desarrollo de la semilla no son necesariamente genes con funciones específicas en la semilla (Tzafrir *et al.*, 2004). En consecuencia, podemos concluir que el acercamiento mutacional, aun siendo importante, no es suficiente para construir un cuadro completo del proceso. La identificación de genes cuya expresión es específica en semilla inmadura puede complementar los acercamientos genéticos.

La generación de colecciones de ESTs y el análisis masivo de expresión mediante micromatrices puede contribuir al esclarecimiento de los procesos que llevan a la formación y maduración de la semilla y del embrión (Ma *et al.*, 2005; Schmid *et al.*, 2005). La secuenciación masiva de cDNAs es una herramienta muy potente no solo como alternativa a la secuenciación completa de genomas, sino también en el estudio de la estructura y de los patrones de expresión de los genes de una especie (Rudd, 2003). Las bases de secuencias contienen un número considerable de ESTs correspondientes a especies de plantas. Por ejemplo, en la versión 200106 (20 de Enero de 2006; <http://www.ncbi.nlm.nih.gov/dbEST/index.html/>) de la dbEST hay depositados 662.884 secuencias de maíz (*Zea mays*), 600.205 de trigo (*Triticum aestivum*), 421.027 de *Arabidopsis thaliana* ó 407.545 de arroz (*Oryza sativa*). Estos números son, sin embargo, pequeños en comparación con las 7.596.977 de secuencias de humanos. El gran número de ESTs obtenidos en humanos ha permitido identificar genes con patrones de expresión interesantes basándose en la frecuencia en que aparecen en las secuencias de genotecas construidas a partir de diferentes órganos, tejidos o condiciones (Bernstein *et al.*, 1996; Bortoluzzi *et al.*, 2000; Vasmatis *et al.*, 1998; Itoh *et al.*, 1998; Miner y Rajkovic, 2003; Huminiecki y Bicknell, 2000; Baranova *et al.*, 2001). Pese a que para *Arabidopsis* la cantidad de ESTs depositadas es mucho más limitada que



para humanos, ya es posible investigar la expresión génica a escala genómica basándose en estos datos.

Otra alternativa al estudio global de la expresión génica en *Arabidopsis* son las micromatrices. Esta tecnología permite el análisis simultáneo de miles de genes en diferentes tejidos, órganos o condiciones. El análisis de los datos obtenidos en estos experimentos nos puede permitir, al menos potencialmente, conocer cuando, donde y en respuesta a que estímulos se expresa cada gen. La existencia de micromatrices conteniendo sondas correspondientes a más de 20.000 genes de *Arabidopsis* (*Arabidopsis* Affymetrix GeneChip®), la accesibilidad pública a los datos de cientos de ensayos (NASC, <http://affymetrix.arabidopsis.info/>; Honys y Twell, 2004; Lloyd y Zakhleniuk, 2004; Schmid *et al.*, 2005; Menges *et al.*, 2005), y la existencia de herramientas para su análisis masivo (Genevestigator® <http://www.genevestigator.ethz.ch/>; Zimmermann *et al.*, 2004, 2005) abren la posibilidad del análisis *in silico* como alternativa a la experimentación directa para la identificación de genes con patrones de expresión concretos en *Arabidopsis*.

#### I.1 Secuenciación de ESTs de semillas inmaduras de *Arabidopsis thaliana*.

La identificación de genes de expresión específica en ciertos estadios de desarrollo basándose en la presencia de ESTs en las bases de secuencias está limitada por el número de secuencias disponibles obtenidas a partir de genotecas de cDNA construidas usando muestras de esos órganos concretos. En el caso de *Arabidopsis*, estas carencias son especialmente evidentes para ciertos órganos y estadios de desarrollo. Por ejemplo, de las más de 420 mil secuencias de EST de *Arabidopsis thaliana* disponibles sólo 10.854 corresponden a semillas inmaduras aisladas y de éstas, únicamente 54 corresponden a estadios tempranos de desarrollo mientras que 10.800 corresponden a estadios intermedios (Figura 12) (White *et al.*, 2000).

Para subsanar parcialmente la carencia de ESTs de semilla inmadura de *Arabidopsis* se construyó una genoteca de cDNAs a partir de semillas de *Arabidopsis* entre los estadios globular intermedio a cotiledón curvado (2 a 6

días después de polinización), se seleccionaron clones cuyo inserto fuera de una longitud mayor a 140 pb y se obtuvo la secuencia parcial de 178 de ellos. La secuenciación fue orientada desde 5' y el tamaño medio de las secuencias obtenidas fue de 579 pb (Cuadros 1 y 2). Estas secuencias de ESTs fueron depositadas en la base de datos de GeneBank bajo los números de acceso AM111128 a AM111305 y el código de la genoteca de cDNAs ATISLA (*Arabidopsis thaliana* Immature Seed Library).

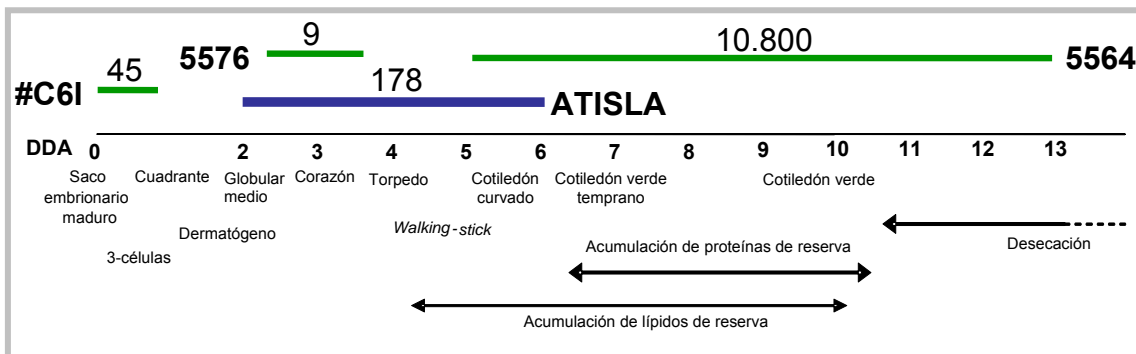


Figura 12.- Distribución de las diferentes genotecas de ESTs de semillas inmaduras de *Arabidopsis thaliana*. Los números sobre las barras horizontales indican el número de ESTs. Los números al costado de barras verdes representan el código de la genoteca en la base de datos de ESTs. La barra azul indica la nueva genoteca obtenida en este trabajo (ATISLA). DDA (0 a 13) corresponde a días después de antripsis y bajo éstos números se señalan los distintos estadios de desarrollo del embrión a medida que transcurren los DDA y los eventos de mayor relevancia durante la formación de la semilla.

Las nuevas secuencias se agruparon en 46 contigs (secuencias con más de un EST) usando el programa CAP3 (<http://fenice.tigem.it/bioprg/interfaces/cap3.html>) (Cuadro 1), quedando 49 singletons (secuencias de un solo EST) (Cuadro 2). El contig más abundante contenía 8 secuencias. Se identificaron, por tanto, un total de 95 genes tentativos. Las secuencias de los genes tentativos fueron comparadas mediante BLAST con el genoma completo de *Arabidopsis* (TAIR; [www.arabidopsis.org](http://www.arabidopsis.org)), identificándose los genes a los que correspondían: 93 genes nucleares y 2 genes cloroplásticos. Para dos de los genes (*At1g60987* y *At2g02490*) aún no se habían depositado ESTs en la base de datos. Las características de los genes se describen en los cuadros 1 (contigs) y 2 (singletons).

La categoría funcional de cada gen se determinó basándose en la información de *Gene Ontology* (GO) en la base de datos TAIR. La función del 30,5 % de los genes es desconocida. Del resto, la categoría más representada corresponde a genes relacionados con la traducción (21,1 % de total). Otras categorías abundantes son el metabolismo de carbohidratos (6,3 %) y el desarrollo (5,3 %). La lista completa se presenta en el cuadro 3.

## 1.2 Selección *in silico* de genes que se expresan específicamente en semillas inmaduras

Después de nuestra aportación, las bases de ESTs contienen 11.032 secuencias provenientes de semillas inmaduras aisladas de *Arabidopsis*. Algunos de ellos corresponderán a genes cuya expresión es únicamente en semilla inmadura. Para seleccionarlos se utilizó un procedimiento de sustracción *in silico* en dos pasos (Figura 13), el primer paso basado en la abundancia de ESTs según el órgano del que se obtuvieron, y el segundo paso basado en el análisis de datos de micromatrices.

Cuadro 1.- Genes con mayor representación en la genoteca de cDNAs ATISLA de semilla inmadura (Contigs).

Código Atg	ESTs genoteca	ESTs total	Definición	Categoría	Localización
At4g12960	8	118	Desconocida	Desconocida	Desconocida
At2g28900	6	69	Proteína canal de membrana	Tráfico subcelular	Mitocondria
At1g08480	4	27	Desconocida	Desconocida	Desconocida
At1g61570	4	11	Translocasa de la membrana interna mitocondrial; subunidad Tim13	Tráfico subcelular	Mitocondria
At1g73550	4	5	Proteína de transferencia de lípidos	Metabolismo de lípidos	Desconocida
At2g38740	4	12	Posible β-fosfoglucomutasa	Metabolismo de carbohidratos	Desconocida
At3g08610	4	34	Desconocida	Desconocida	Desconocida
At4g00585	4	25	Desconocida	Desconocida	Desconocida
At5g54800	4	25	Translocador de glucosa 6 fosfato/fosfato	Metabolismo de carbohidratos	Desconocida
At5g56030	4	137	Proteína <i>Heat shock</i> 81-2	Procesamiento de proteínas	Citoplasma
AtCg00480	4	120	Cadena β ATP-sintasa	Energía	Mitocondria
At1g11840	3	152	Lactoilglutación liasa	Metabolismo de carbohidratos	Desconocida
At1g26470	3	8	Desconocida	Desconocida	Desconocida
At1g78380	3	138	Glutación transferasa	Detoxificación de oxígeno	Citoplasma
At3g28500	3	32	Proteína ribosomal ácida 60S, P2	Traducción	Citoplasma
At4g12870	3	10	Desconocida	Desconocida	Desconocida
At4g32320	3	12	L-ascorbato peroxidasa	Detoxificación de oxígeno	Desconocida
At5g20165	3	7	Desconocida	Desconocida	Sistema de endomembrana
At5g23440	3	8	Proteína tipo ácido lipóico sintasa	Fotosíntesis	Cloroplasto
At5g64310	3	49	Proteína arabinogalactano	Desarrollo	Membrana
At1g05190	2	43	Posible proteína ribosomal 50S de cloroplasto, L6	Traducción	Cloroplasto
At1g12000	2	66	Subunidad β de la fosfofructoquinasa dependiente de pirofosfato	Metabolismo de carbohidratos	Citoplasma
At1g21190	2	5	Proteína tipo Sm	Transcripción, splicing	Núcleo
At1g41880	2	48	Proteína ribosomal 60S, L35a	Traducción	Citoplasma
At1g54340	2	7	Isocitrato dehidrogenasa específica de NADP	Metabolismo de carbohidratos	Desconocida
At1g60987	2	2	Desconocida	Desconocida	Desconocida
At1g76200	2	37	<i>brassinolide enhanced gene (BLE1)</i>	Desarrollo	Desconocida
At2g02490	2	2	Desconocida	Desconocida	Desconocida
At2g21185	2	11	Desconocida	Desconocida	Desconocida
At2g38900	2	6	Inhibidor de proteasas	Defensa	Desconocida
At3g02310	2	8	SEPALLATA2 (Proteína MADS box tipo Agamous, AGL4)	Regulación de expresión de genes	Núcleo
At3g05910	2	31	Posible pectinacetiltransferasa	Desarrollo	Pared celular
At3g08770	2	47	Precursor de la proteína 6 de transferencia de lípidos no específica (LTP6)	Metabolismo de lípidos	Desconocida
At3g26060	2	35	Proteína peroxirredoxina tipo Q	Detoxificación de oxígeno	Desconocida
At3g27830	2	70	Proteína ribosomal 50S, L12-1; precursor cloroplasto	Traducción	Cloroplasto
At4g39880	2	12	Desconocida	Desconocida	Desconocida
At5g05270	2	19	Similar a chalcona-flavonona isomerasa	Metabolismo secundario	Desconocida
At5g13410	2	7	Proteína Tilacoide lumenal 18kDa; precursor cloroplasto	Desconocida	Cloroplasto
At5g26667	2	20	Uridilato quinasa (UK) (Uridina monofosfato quinasa)(UMP quinasa)(UMP/CMP quinasa)	Regulación de expresión de genes	Desconocida
At5g27520	2	20	Desconocida	Desconocida	Desconocida
At5g37510	2	59	Subunidad de 75kDa de la NADH-ubiquinona oxidoreductasa; precursor mitocondrial (Complejo I-75Kd)	Metabolismo de aminoácidos	Mitocondria
At5g42080	2	29	Proteína de unión a GTP tipo dinamin	Ciclo de división celular	Citoplasma
At5g45390	2	56	Proteína Clp tipo proteasa dependiente de ATP	Procesamiento de proteínas	Cloroplasto
At5g48480	2	21	Desconocida	Desconocida	Desconocida
At5g62050	2	16	Precursor mitocondrial de la proteína de membrana interna OXA1 (Proteína 1 de unión oxidasa)(AtOXA1)	Procesamiento de proteínas	Mitocondria
At5g64900	2	9	Desconocida	Desconocida	Desconocida

Cuadro 2.- Genes con un solo EST (Singletons) identificados en la genoteca de cDNAs ATISLA de semilla inmadura.

Código Atg	ESTs genoteca	ESTs total	Definición	Categoría	Localización
At1g09200	1	24	Proteína histona H3.2	Ciclo de división celular	Núcleo
At1g15040	1	5	Desconocida	Desconocida	Desconocida
At1g18460	1	11	Posible lipasa	Metabolismo de lípidos	Desconocida
At1g23490	1	114	Factor tipo 1 de ribosilación de ADP	Procesamiento de proteínas	Citoplasma
At1g25260	1	26	Proteína ribosomal L10	Traducción	Citoplasma
At1g29040	1	15	Desconocida	Desconocida	Desconocida
At1g43890	1	24	ATRab18	Regulación de expresión de genes	Desconocida
At1g56700	1	33	Proteína tipo pirrolodina carboxil-peptidasa	Degradación de proteínas	Desconocida
At1g62060	1	61	Desconocida	Desconocida	Desconocida
At1g70190	1	9	Posible proteína ribosomal 50S, L12	Traducción	Citoplasma
At1g70830	1	127	Proteína 28 tipo MLP	Metabolismo secundario	Desconocida
At1g74340	1	4	Posible Dolicol-fosfato ( $\beta$ -D) manosiltransferasa 2	Procesamiento de proteínas	Membrana del R. endoplásmico
At1g75220	1	23	Similar a proteína integral de membrana	Metabolismo de carbohidratos	Membrana plasmática
At1g80830	1	17	Proteína transportadora de metales, Nramp1	Tráfico subcelular	Desconocida
At2g01250	1	77	Proteína ribosomal 60S, L7-1	Traducción	Citoplasma
At2g05790	1	8	Proteína tipo $\beta$ -1,3-glucanasa	Defensa	Desconocida
At2g27260	1	9	Desconocida	Desconocida	Desconocida
At2g32980	1	9	Desconocida	Desconocida	Desconocida
At2g34700	1	6	Glicoproteína rica en prolina	Desarrollo	Pared celular
At2g38670	1	21	Probable fosfolípido citidiltransferasa	Metabolismo de lípidos	Desconocida
At3g01130	1	5	Proteínas ancladas a GPI	Desconocida	Desconocida
At3g06680	1	46	Proteína ribosomal L29	Traducción	Citoplasma
At3g10090	1	32	Proteína ribosomal S28	Traducción	Citoplasma
At3g11940	1	116	Proteína ribosomal 40S, S5-2	Traducción	Citoplasma
At3g20210	1	63	Enzima de procesamiento vacuolar (proteínasa)	Degradación de proteínas	Vacuola
At3g22230	1	39	Proteína ribosomal 60S, L27	Traducción	Citoplasma
At3g24100	1	30	Desconocida	Desconocida	Desconocida
At3g48930	1	60	Proteína ribosomal 40S, S11-1	Traducción	Citoplasma
At3g53740	1	85	Proteína ribosomal 60S, L36-2	Traducción	Citoplasma
At3g60245	1	99	Proteína ribosomal 60S, L37a	Traducción	Citoplasma
At4g00100	1	55	Proteína ribosomal 40S, S13-2	Traducción	Citoplasma
At4g09090	1	3	Proteína tipo $\beta$ -1,3-glucanasa	Defensa	Desconocida
At4g18100	1	57	Proteína ribosomal 60S, L32A	Traducción	Citoplasma
At4g23630	1	90	Desconocida	Desconocida	Desconocida
At4g32680	1	11	Desconocida	Desconocida	Desconocida
At4g33865	1	33	Proteína tipo proteína ribosomal S29	Traducción	Citoplasma
At4g37830	1	37	Posible subunidad V1a de la citocromo c oxidasa	Desconocida	Desconocida
At5g13930	1	150	chalcona sintasa	Metabolismo secundario	Desconocida
At5g19900	1	6	Factor A de interacción PRL1	Regulación de expresión de genes	Cloroplasto
At5g25770	1	20	Desconocida	Desconocida	Desconocida
At5g27950	1	7	Proteína 5 tipo kinesina BY-2	Ciclo de división celular	Citoesqueleto
At5g42020	1	56	Precursor de la proteína 2 de unión luminal (Bip2)(AtBP2)	Ciclo de división celular	Desconocida
At5g54770	1	241	Thiazole biosynthetic enzyme, Cloroplasto precursor (ARA6)	Metabolismo de ADN	Cloroplasto
At5g55250	1	4	S-adenosil-L-metionina: proteína tipo ácido salicílico carboxil metiltransferasa	Defensa	Desconocida
At5g56940	1	10	Proteína ribosomal 30S, S16	Traducción	Citoplasma
At5g58130	1	16	Desconocida	Desconocida	Desconocida
At5g59880	1	34	Factor 3 depolimerizante de actina (ADF 3)	Desarrollo	Citoesqueleto
At5g64130	1	53	Desconocida	Desconocida	Desconocida
orf 146-ct	1	33	Proteína ribosomal S12	Traducción	Cloroplasto

Cuadro 3.- Categorías funcionales de las ESTs secuenciados de ATISLA

Categoría funcional	Número de genes	% de genes
Ciclo de división celular	4	4,2
Defensa	4	4,2
Degradación de proteínas	3	3,2
Desarrollo	5	5,3
Metabolismo de las especies de oxígeno reactivo	3	3,2
Energía	1	1,1
Fotosíntesis	1	1,1
Metabolismo de ácidos nucleicos	1	1,1
Metabolismo de aminoácidos	1	1,1
Metabolismo de carbohidratos	6	6,3
Metabolismo de lípidos	4	4,2
Metabolismo secundario	3	3,2
Procesamiento de proteínas	4	4,2
Regulación de la transcripción	4	4,2
Traducción	20	21,1
Tráfico subcelular y transporte	3	3,2
Transcripción, <i>splicing</i>	1	1,1
Desconocida	29	30,5

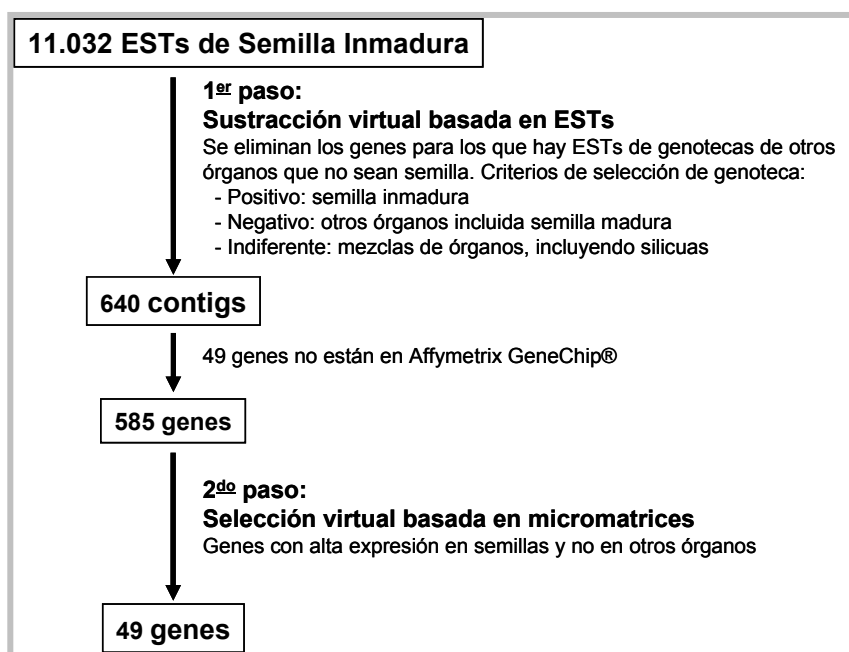


Figura 13.- Esquema de procedimiento de selección *in silico* para identificar genes expresados específicamente en semillas inmaduras de *Arabidopsis thaliana*.

El objetivo del primer paso fue identificar genes para los cuales se hubieran secuenciado ESTs en genotecas de semilla inmadura pero no de genotecas construidas a partir de otros órganos. Las genotecas de ESTs de *Arabidopsis* (TIGR *Arabidopsis* Gene Index, [http://www.tigr.org/tigr-scripts/tgi/T\\_index.cgi?species=arab](http://www.tigr.org/tigr-scripts/tgi/T_index.cgi?species=arab)) se dividieron en tres categorías de

acuerdo a los órganos a partir de los cuales habían sido construidas (Cuadro 4):

- a) Semillas inmaduras: incluye 11.032 ESTs de cuatro genotecas de cDNA.
- b) Otros tejidos: incluye 50.992 ESTs de 78 genotecas de cDNA obtenidas de tejidos vegetativos, flores sin polinizar y semillas secas. Los genes que tenían ESTs procedentes de estas genotecas fueron eliminados.
- c) No informativas: aquéllas genotecas obtenidas a partir de mezclas de órganos o de plantas completas, incluyendo genotecas de silículas.

Cuadro 4.- Clasificación de las genotecas de *Arabidopsis thaliana* de acuerdo al órgano del cual se obtuvieron:

<b>Genotecas de semillas inmaduras.</b>			
Código	Número de ESTs	Número de TCs <sup>1</sup>	Descripción de la genoteca
5564	10800	3939	5 – 13 días después de floración
5576	9	9	Estado de corazón
#C6I	45	38	Óvulos
ATISLA	178	95	2 – 6 días después de antesis
<b>Genotecas de otros tejidos.</b>			
<i>Semillas secas</i>			
NH14	346	183	Perp-dry-seedA; semillas secas
<i>Hojas y tallos</i>			
#EP3	84	83	Tejido vegetativo aéreo de 4 semanas
#F0G	2416	1283	Tejido vegetativo adulto; Columbia 0
#DGL	5419	2115	Hoja senescente
2338	611	581	Roseta 1; Columbia 0
2340	874	844	Roseta 2; Columbia 0
2341	785	760	Roseta 3; Columbia 0
NH29	1	1	Parte aérea; 2 semanas
5335	12264	3654	Órganos aéreos; 2 – 6 semanas
3792	17	17	Órganos aéreos; 2 – 6 semanas; Columbia 0
4063	3348	1895	Órganos aéreos; 2 – 6 semanas
#A5R	17	14	RAFL12; roseta
#A5S	64	55	RAFL13; roseta
4921	642	374	AB; hoja de plántula
NH25	453	346	CD4-14; hipocotilo de plántulas
NH26	819	585	CD4-15; hipocotilo de plántulas
NH27	563	400	CD4-16; hipocotilo de plántulas
NH28	724	531	CD4-13; hipocotilo de plántulas
4932	1007	604	AA; hoja
#BU2	5	5	Roseta
NH12	322	187	Ors-A; parte aérea
NH16	667	408	Strasbourg-A; hoja
NH17	37	28	Strasbourg-FE; hoja
<i>Raíz</i>			
#DOU	4	2	Raíces
2336	538	504	Raíz 1; Columbia Col – 0
2337	645	618	Raíz 2; Columbia Col – 0
5336	17574	5266	Raíces; Columbia
#A5T	24481	4731	RAFL14; raíz
#BMJ	20	4	Raíz
#E3R	1434	955	MPIZ-ADIS-066; raíz
NH13	239	166	Ors-B; raíz

Cuadro 4.- Clasificación de las genotecas de *Arabidopsis thaliana* de acuerdo al órgano del cual se obtuvieron (continuación):

Código	Número de ESTs	Número de TCs <sup>1</sup>	Descripción de la genoteca
<i>Flor</i>			
#F0J	1724	772	Flor y botones florales
2334	356	342	Inflorescencia-1; Columbia Col-0
2335	674	650	Inflorescencia-2; Columbia Col-0
5337	5719	2514	Botones florales; Columbia
NH08	2	2	Grenoble-A; botones florales
NH09	984	525	Grenoble-B; botones florales
#E55	1447	1083	ag-1 35S:AG-GR; inflorescencia
#EA9	4382	1984	Meristema de inflorescencia
NH36	201	177	Inflorescencia
NH35	20	8	Inflorescencia
#CCH	1165	774	MPIZ-ADIS-035; inflorescencia
<i>Estrés biótico</i>			
#BKG	3301	1796	Hoja infectada
#CAH	232	203	<i>Pseudomonas syringae</i> avirulenta
#CAI	140	121	<i>Pseudomonas syringae</i> virulenta
#C6J	243	212	<i>Peronospora parasitica</i> virulenta
#C6K	110	104	<i>Peronospora parasitica</i> avirulenta
#C6L	11	11	<i>Peronospora parasitica</i> virulenta
#C6M	2	1	<i>Peronospora parasitica</i> avirulenta
1725	22	16	Raíz infectada por nemátodos
NH15	121	73	Ra147.1; células patógeno
<i>Estrés abiótico</i>			
#5GJ	170	138	Ozono
#5GK	27	25	Ozono
#5GL	128	107	Ozono
6523	148	125	Sal
6524	536	224	Sal
#C6P	142	130	Ozono
<i>Otras</i>			
#CFQ	80	64	Productos de RT-PCR clonados; plántula
2370	16	16	Semillas en germinación
4924	333	199	AC; mezcla de hoja y raíz de plántula
5338	1135	747	Cultivo líquido de plántulas; Columbia
NH18	1503	906	Versalles-VB; plántula
NH19	124	113	Versalles-VC; plántula
NH20	7	6	Versalles-VD; plántula
#A64	433	225	RAFL2; roseta
#A65	384	200	RAFL3; roseta
#A66	1885	999	RAFL4; roseta
#A67	3767	1911	RAFL5; roseta
#A69	3494	1650	RAFL7; roseta
#A6A	3373	1650	RAFL8; roseta
#C6N	3	3	Metiljasmonato
#C6H	56	52	Metiljasmonato
NH01	123	75	AC13D; suspensión celular
NH02	897	632	AC16H; células
#CAG	107	77	Ácido salicílico
NH30	27	17	Suspensión celular
#F0H	2308	915	Callos de Col-0 tratados con hormonas
#CCG	2390	1219	MPIZ-ADIS-008; plántula
<b>Genotecas No informativas</b>			
<i>Silicua</i>			
#F01	919	445	Silicua; Col-0
2369	4	4	Silicua inmadura
5339	12591	4879	Silicua verde; Columbia
NH05	389	194	Gif-SeedA; silicua verde
NH06	284	210	Gif-SeedA+B; silicua verde
NH07	116	65	Gif-SiliqueB; silicua verde
<i>Mezcla</i>			
NH37	36	33	Clon CIC7E11 (YAC) de región específica de cDNA
NH38	74	63	Clon CIC8B11 (YAC) de región específica de cDNA
#A5U	14386	4654	RAFL15; flor y silicua
#BLJ	4	3	De Veylder L.
NH39	47	40	pSMASH; planta completa
NH10	154	138	Lambda-PRL1; mezcla
NH11	27641	7771	Lambda-PRL2; mezcla
#CCC	1782	953	MPIZ-ADIS-027; mezcla
#CCD	1554	874	MPIZ-ADIS-014; mezcla
#CCE	1611	951	MPIZ-ADIS-013; mezcla
#CCF	1080	722	MPIZ-ADIS-012; mezcla



Cuadro 4.- Clasificación de las genotecas de *Arabidopsis thaliana* de acuerdo al órgano del cual se obtuvieron (continuación):

Código	Número de ESTs	Número de TCs <sup>1</sup>	Descripción de la genoteca
<i>Mezcla (continuación)</i>			
2342	245	222	Mezcla
#A62	25819	5766	RAFL19; flor y silicua
#A68	8023	3297	RAFL6; mezcla
#A6B	29153	7180	RAFL9; mezcla
#A6O	2645	1399	RAFL11; mezcla
NH03	99	70	AT-NHC; planta completa
NH44	176	114	Planta completa
#FGG	280	78	Productos de RT-PCR (CSHL); planta completa
<i>Otros</i>			
NH04	3	3	M.W. Schena
NH43	2	2	Lambda-PRL2 (P. Kapranov)
NH41	6	6	mRNA (J. Pleck)
#A60	14446	4697	RAFL17
#A61	1250	843	RAFL18
#A63	12784	4242	RAFL21
2339	3014	2559	Clones; Ohio State
NH31	34	26	RIKEN-PMB-FL1
2373	44	33	Lambda ZipLox designated PRL2
NH40	59	36	Motohashi, R.
NH34	1	0	Tan, K.
NH32	5	3	Josefsson, L.G.
JnAr	22505	18035	Modelo de gen Ath1
Ceres	5017	5005	cDNA completos de CERES
7052	21	20	FL5
7053	5	5	FL3
7054	6	6	FL2
7055	1	1	FL6
2741	5	1	Columbia
#A5V	26509	6333	RAFL16

1: TC, *tentative contig* de acuerdo a TIGR.

El primer criterio de sustracción fue superado por 640 contigs. 2 de ellos correspondían a genes de cloroplastos, 3 correspondían a genes de mitocondrias y 26 tenían homología con partes del genoma de *Arabidopsis* en las que no se habían definido genes y podían corresponder a elementos transponibles. Al menos 2 de ellos tenían similitud con transposones conocidos.

El segundo paso de selección estaba basado en los resultados de las hibridaciones de micromatrices contenidos en la base de datos de Affymetrix GeneChip® de *Arabidopsis* disponible en el sitio de análisis Geneinvestigator® (<http://www.geneinvestigator.ethz.edu>) (Zimmermann *et al.*, 2004, 2005). Se utilizó el programa Meta-Analyzer que realiza para cada uno de los genes un mapa de valores normalizados de intensidad de señal que corresponde a los diferentes órganos de la planta. Los rangos de valores van de 0 a 100, siendo 100 el valor más alto de hibridación. Los genes se seleccionaron utilizando los siguientes criterios:

- i) La expresión en semillas debía ser mayor que 80.

- ii) La expresión en los otros órganos debía ser menor que cinco, excepto en carpelos e inflorescencias, ya que estos dos órganos podrían contener semillas inmaduras a estadíos muy tempranos después de polinización, y en silículas.
- iii) Los niveles de expresión en semillas debían ser mayores o iguales a la expresión en silículas, carpelos o inflorescencias.

De los 634 genes seleccionados (Ver Apéndice II), 49 no fueron considerados en el segundo análisis debido a que no están incluidos en el Affymetrix 22 K GeneChip® de Arabidopsis. De los 585 genes restantes 49 (8%) cumplieron los criterios de selección y pueden representar a genes que se expresan específicamente en semillas inmaduras (Cuadro 5).

Cuadro 5.- Genes seleccionados por la doble sustracción *in silico*.

Código AGI	ESTs semilla inmadura	ESTs indiferentes	Definición	Patrón de expresión <sup>1</sup>	Mutantes	Grupos en tándem	Duplicación de segmentos cromosómicos
At1g03790	1	5	Proteína de la familia Zinc finger (tipo CCCH)	llc	-	1	1
At1g03890	41	22	Cruciferina, proteína 12S de almacenamiento de la semilla	llb	-	2	1
At1g14950	2	15	Proteína mayoritaria en látex	llc	-	4	1
At1g48130	2	12	Peroxirredoxina	llb	-	1	1
At1g48660	1	0	Proteína de la familia de respuesta a auxina GH3	llc	-	3	1
At1g62060	32	25	Desconocida	lla	-	2	1
At1g65090	2	6	Desconocida	llb	-	1	1
At1g67100	3	3	Proteína específica de la semilla Bn15D17A	llb	-	1	1
At1g73190	8	16	Proteína intrínseca del tonoplasto 3.1	llb	-	1	2
At1g80090	3	2	Desconocida	llb	-	1	1
At2g28420	1	4	Proteína de la familia lactoilglutación liasa	llc	-	1	1
At2g33520	1	1	Proteína expresada	llc	-	1	1
At2g34700	2	4	Glicoproteína rica en prolina	l	-	1	1
At3g01570	15	52	Oleosina	llb	-	1	1
At3g04170	1	0	Proteína tipo Germina de la subfamilia 1	l	-	5	1
At3g04190	1	0	Proteína tipo Germin de la subfamilia 1	l	-	5	1
At3g12960	1	0	Similar a proteína PM28 de maduración de semilla	llc	-	1	1
At3g24650	6	3	Proteína ABI3	llb	Abi3 <sup>2</sup>	1	1
At3g27660	7	0	Oleosina	llb	-	1	1
At3g48580	1	1	Transferasa xiloglucan : xiloglucosil	llc	-	1	1

Cuadro 5.- Genes seleccionados por la doble sustracción *in silico* (continuación).

Código genético AGI	ESTs semilla inmadura	ESTs indiferentes	Definición	Patrón de expresión <sup>1</sup>	Mutantes	Grupos en tándem	Duplicación de segmentos
At3g54940	4	18	Cisteín-proteinasa	llb	-	1	1
At3g60730	2	2	Proteína tipo pectinesterasa	llc	-	1	1
At3g61040	1	0	Citocromo P450 tipo monoxigenasa	llc	-	1	1
At3g62730	55	17	Proteína relacionada con desecación	llb	-	1	1
At3g63040	1	0	Proteína expresada	llb	-	1	1
At4g25140	1	5	Oleosina/Proteína rica en glicina	llb	-	1	1
At4g27150	68	33	Precursor 2S de proteína 2 de almacenamiento en semillas	llb	-	4	1
At4g28520	92	4	Proteína 12S de almacenamiento en semilla, cruciferina (CRU3)	llb	-	1	1
At4g36700	48	16	Proteína tipo globulina	lla	-	1	1
At4g37050	2	5	Patatina	lla	-	3	1
At5g01670	1	1	Proteína tipo aldosa reductasa	llc	-	1	1
At5g03860	1	18	Malato sintasa	llc	-	1	1
At5g04010	1	0	Proteína expresada	llc	-	1	1
At5g07190	10	15	Proteína 3 específica de embrión (ATS3)	llb	-	1	1
At5g09640	10	4	Serina carboxipeptidasa	I	Sng2 <sup>3</sup>	1	1
At5g22470	8	1	Proteína de la familia polimerasa Poli (ADP-ribosa)	llc	-	1	1
At5g40420	39	68	Oleosina	llb	-	1	1
At5g44310	5	1	Proteína abundante en embriogénesis tardía	llc	-	1	1
At5g45690	4	6	Desconocida	llc	-	1	1
At5g45830	1	1	Proteína expresada	llc	-	1	1
At5g48100	30	19	Lacasa	lla	-	1	1
At5g49190	9	0	Sacarosa sintasa (SUS2)	I	-	1	1
At5g50700	9	41	11-β-hidroxiesteroide dehidrogenasa	llb	-	2	1
At5g54740	7	37	Proteína 2S de reserva	llb	-	1	1
At5g55240	6	3	Proteína 1 específica de embrión	llb	-	1	1
At5g57260	1	0	Citocromo P450	llb	-	1	2
At5g59170	11	6	Precursor de proteínas de pared celular / extensina	llb	-	1	1
At5g62490	2	5	ATHVA22b	llc	-	1	1
At5g62800	1	0	Proteína de la familia <i>Seven in absentia</i> (SINA)	llb	-	1	1

<sup>(1)</sup>: I, genes de elevada expresión entre estadios de torpedo tardío y *walking-stick*; II, genes de elevada expresión entre cotiledón temprano y tardío (subcategorías: Ila, expresión muy temprana; llb, expresión temprana; llc, expresión de estadio intermedio) (Ver punto I.5)

<sup>(2)</sup>: Mutante insensible a ácido abscísico y que pierde la latencia de la semilla

<sup>(3)</sup>: Mutante que acumula sinapoilglucosa en lugar de sinapoilcolina

La presencia entre los seleccionados de varios genes previamente caracterizados que se expresan de manera específica en semilla demuestra la bondad del método de selección. Por ejemplo, se encuentran entre los genes seleccionados:

- *abi3*, un factor de transcripción que regula las respuestas de ácido abscísico durante el desarrollo de la semilla (Giraudat *et al.*, 1992).
- *At1g48130*, que codifica una peroxirredoxina (PER1) y su expresión esta limitada al tejido embrionario (Haslekas *et al.*, 1998).
- *At1g67100*, un homólogo al gen *Bn15D17A* de colza (*Brassica napus*) que tiene una elevada y específica expresión en embrión y en la cubierta en los estadios tempranos del desarrollo de la semilla (Dong *et al.*, 2004).
- *At5g07190* y *At5g55240*, que se expresan específicamente en embrión y fueron obtenidas mediante un experimento de *differential display* (Nuccio y Thomas, 1999) (Cuadro 5).

A la vista de los resultados cabría preguntarse si el primer paso de selección (ESTs) es necesario, o si bastaría con el uso de los datos de micromatrices. Por ello decidimos realizar una estimación de lo que ocurriría si aplicáramos únicamente la sustracción basada en micromatrices, con los mismos criterios, pero sin selección previa. Se analizaron los primeros 1500 genes del cromosoma I de acuerdo con el código AGI presentes en el Affymetrix GeneChip® (*At1g01010* al *At1g18340*). 28 de los 1500 genes pasaron la selección, lo cual representa un 1,9%. Si asumimos que todos los genes de *Arabidopsis* guardan la misma proporción, un total de 550 genes pasarían esta selección. Este número es mucho mayor que el obtenido en la doble selección, pero incluye genes para los cuales se han secuenciado ESTs a partir de genotecas construidas con RNAs de órganos diferentes a semillas. De los 28 genes seleccionados:

- 7 también pasaron la sustracción basada en ESTs (25 %).
- 21 no pasaron el criterio de selección por ESTs (75 %):
  - 4 no fueron seleccionados por no tener ESTs (14 %).
  - 5 no tienen ESTs de semilla inmadura (18 %).
  - 7 tienen ESTs en otros órganos que no son semilla inmadura (25 %).

Vemos por tanto que la doble sustracción permite una selección más precisa de los genes.

### I.3 Validación experimental de los patrones de expresión de genes seleccionados.

La calidad del sistema de sustracción fue puesta a prueba mediante el análisis experimental de los patrones de expresión de algunos de los genes de los cuales no se tenían datos previos sobre su expresión en la literatura. Se utilizó la técnica de la RT-PCR semicuantitativa (Figura 14).

Se seleccionaron 10 genes, cinco de los cuales habían superado únicamente la sustracción basada en ESTs y no la de micromatrices. Los otros cinco habían superado los dos criterios de selección. Se agregaron dos genes utilizados como controles: el de actina, que se expresa en todos los tejidos y que se utilizó como control de carga, y *AtEm6*, que se expresa específicamente durante la embriogénesis tardía (Vicient *et al.*, 2000). Se realizaron ensayos de amplificación a partir de diversos tejidos vegetativos y de silicuas a diferentes estadios de desarrollo. Para los 10 genes se observaron los más altos niveles de expresión en silicuas. En algunos genes se observan amplificaciones más débiles en otros órganos, si bien es más frecuente entre los genes seleccionados mediante sustracción basada únicamente en ESTs. Estos resultados demuestran que el sistema de selección *in silico* que hemos utilizado produce los resultados esperados. Por otro lado, podemos observar que la combinación de selección por ESTs más selección por micromatrices produce mejores resultados que la selección únicamente por ESTs.

La expresión específica en semillas se demostró además mediante hibridación *in situ* para el gen *At5g22470* que codifica una polimerasa poliADP-ribosa (PARP) (Figura 15) y para el gen *abi3*, un factor de transcripción que regula las respuestas de ácido abscísico durante el desarrollo de la semilla (Giraudat *et al.*, 1992). Los transcritos de *At5g22470* aparecen únicamente en el embrión y no en el endospermo, pericarpo, valvas o septum. El perfil de expresión del gen *At5g22470* es consistente con la transcripción específica predicha en semilla. La hibridación de la sonda antisentido del gen *abi3* se detectó específicamente en embrión de *Arabidopsis thaliana* (Figura 16).

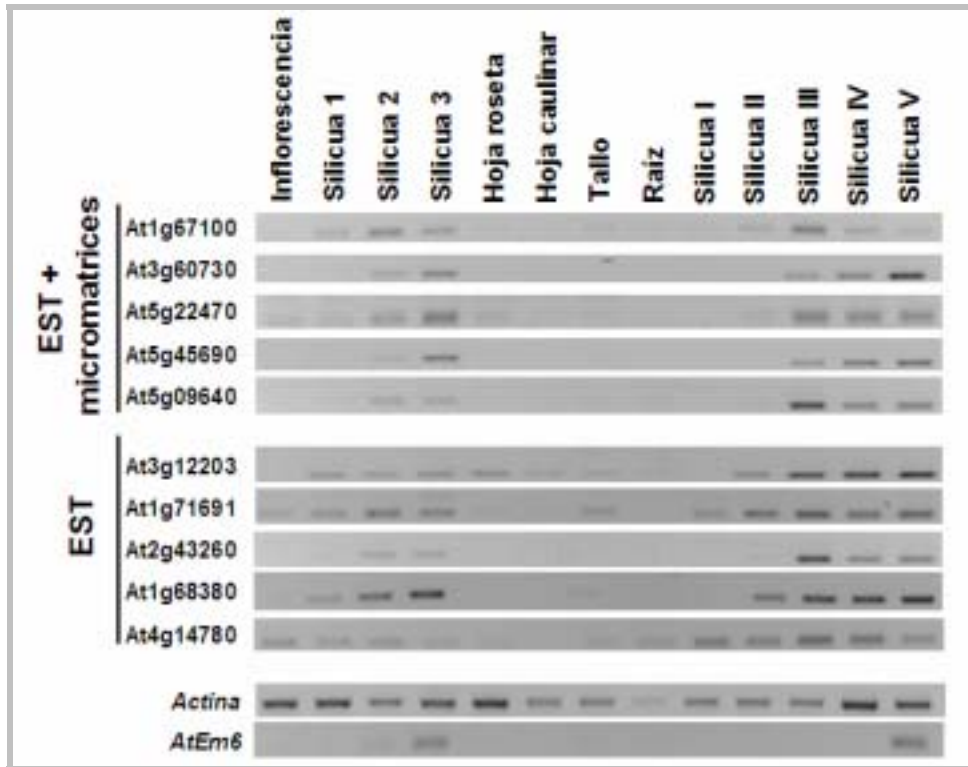


Figura 14.- Análisis de RT-PCR semicuantitativa del patrón de expresión de 10 genes seleccionados *in silico*. “EST + micromatrices”, indican los genes aislados por la combinación de la colección de ESTs y el análisis de datos de micromatrices. “EST” indica los genes aislados solo por la selección de ESTs. Silicuas 1 a 3 corresponden a silicuas a diferentes estados de desarrollo (1, silicuas verdes jóvenes; 2, desarrolladas completamente; 3, silicuas en proceso de desecación. Silicuas I a V, corresponden a silicuas en diferentes estados de desarrollo (I, 0-4 Días Después de Antesis (DDA); II, 4-8 DDA; III, 8-12 DDA; IV, 12-16 DDA y V, 17-21 DDA. *Actina* y *AtEm6*, controles.

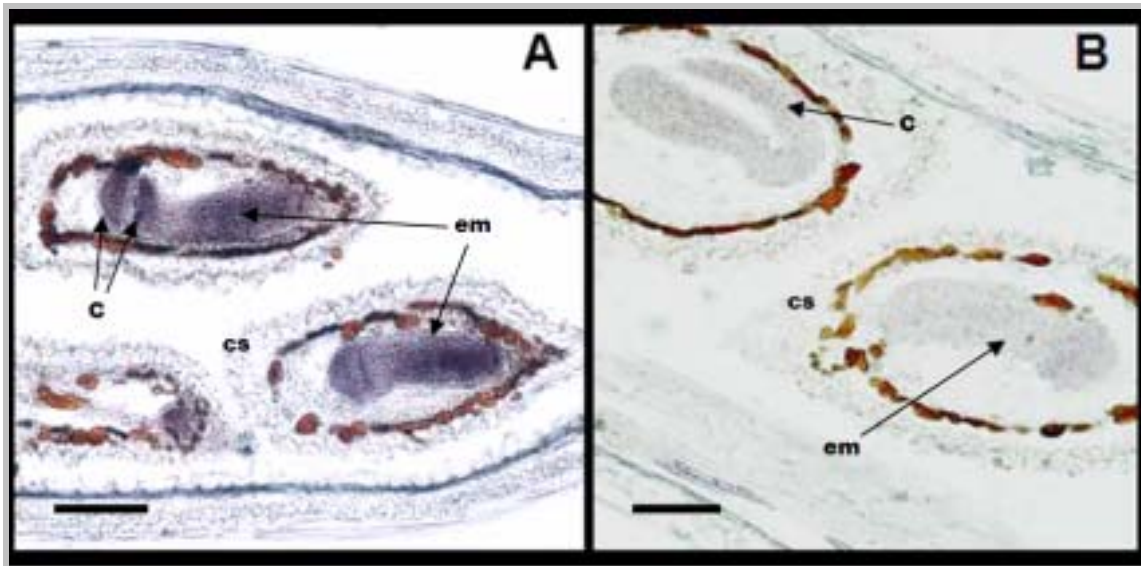


Figura 15.- Análisis de hibridación *in situ* del gen expresado específicamente en semilla. Marcaje de transcritos específicos de semilla, en embriones en estadio de torpedo tardío. Cortes transversales de silicuas de *Arabidopsis* con RNA marcado con digoxigenina del gen *At5g22470* y observado bajo microscopía óptica. A, marcaje de cadena antisentido; B, marcaje de cadena con sentido. Abreviaturas: C, cotiledón; em, embrión; cs, cubierta seminal. Barras, 50µm.

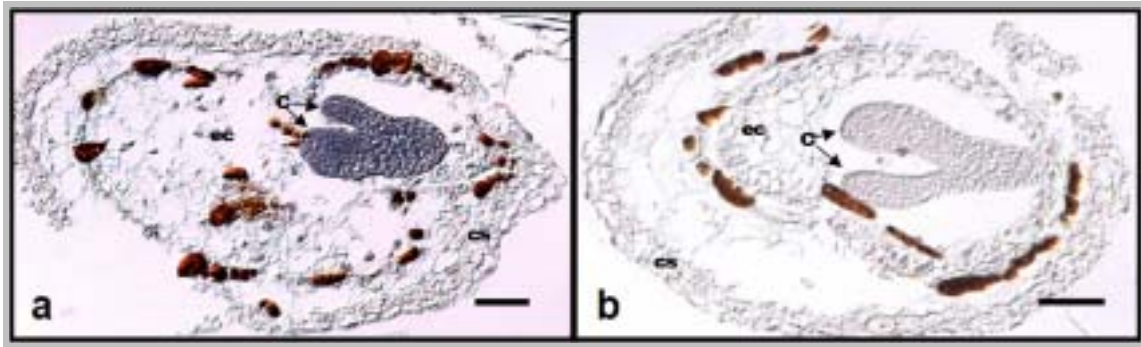


Figura 16.- Análisis de hibridación *in situ* del gen *abi3* en silicuas inmaduras de *Arabidopsis thaliana*. (a) Corresponde a la hibridación de la sonda antisentido, (b) a la hibridación de la sonda sentido. Abreviaturas: c, cotiledones; ec, endospermo celularizado; cs, cubierta seminal. Barras, 50  $\mu$ m.

#### I.4 Clasificación funcional de los genes seleccionados

Los genes seleccionados de expresión específica en semilla se agruparon en diferentes categorías funcionales basándose en *Gene Ontology* (GO) *Consortium* a través del consorcio de información de *Arabidopsis* ([www.arabidopsis.org](http://www.arabidopsis.org)) (Cuadro 6). 35 genes (71,4 %) fueron asignados a ocho categorías diferentes y no se pudo determinar la función de 14 (28,6%), que fueron agrupados en la categoría “desconocida”. Este porcentaje es menor que el observado para el genoma completo (Berardini *et al.*, 2004) (aunque no significativamente,  $\alpha = 0.05$ ). La identificación de 14 genes de función desconocida en la expresión específica de semilla es de interés, particularmente *At1g62060*, que está representado por un total de 57 secuencias de ESTs en la base de datos (Cuadro 5). Otros dos de los genes codifican proteínas de tipo germina (*At3g04170* y *At3g04190*) y cuatro más son genes de expresión específica en embrión pero de función desconocida (*At1g67100*, *At3g12960*, *At5g07190* y *At5g55240*).

Los genes involucrados en la reserva de nutrientes suponen el 20,4% de la selección e incluyen cuatro genes que codifican oleosinas, tres globulinas, dos cruciferinas y una patatina. La abundancia de este grupo es significativamente mayor que el total del genoma (0,2 %) (Berardini *et al.*, 2004). Esta categoría incluye los genes con un mayor número de ESTs en las bases de secuencias, lo cual indica que sus niveles de expresión son muy altos.

La tercera categoría más abundante es la de respuesta a estrés abiótico. Incluye seis genes (12,2 %) y es significativamente más abundante que el total del genoma (3,1 %). Tres de los genes codifican enzimas relacionados con el estrés oxidativo (*At1g48130*, *At5g48100* y *At5g50700*), dos están relacionados con desecación (*At5g62490* y *At5g44310*) y uno es un gen inducible por ABA (*At3g62730*). Esta categoría es la segunda en abundancia de ESTs por gen.

Cuadro 6.- Categorías funcionales

Categoría funcional	% genes		Nº de ESTs medio por gen
	Genoma completo <sup>1</sup>	Sustraídos (p-value) <sup>2</sup>	
Reserva de nutrientes	0,2	20,4 <sup>0,00*</sup>	56,2 ± 12,3
Ciclo de división celular	2,3	0,0 <sup>0,63</sup>	-
Defensa	0,9	0,0 <sup>1,00</sup>	-
Desarrollo	6,0	8,2 <sup>0,54</sup>	7,0 ± 3,5
Energía	4,0	4,1 <sup>1,00</sup>	1,0 ± 0,0
Fotosíntesis	0,3	0,0 <sup>1,00</sup>	-
Metabolismo	6,4	0,0 <sup>0,07</sup>	-
Metabolismo de ácidos nucleicos	3,1	0,0 <sup>0,41</sup>	-
Metabolismo de aminoácidos	0,1	0,0 <sup>1,00</sup>	-
Metabolismo de carbohidratos	2,4	10,2 <sup>0,01*</sup>	7,4 ± 3,2
Metabolismo de lípidos	0,9	0,0 <sup>1,00</sup>	-
Metabolismo secundario	0,7	2,0 <sup>0,28</sup>	17
Procesamiento de proteínas	9,4	10,2 <sup>0,81</sup>	9,4 ± 4,2
Regulación de la transcripción	7,4	4,1 <sup>0,58</sup>	7,5 ± 2,9
Respuesta a estrés abiótico	3,1	12,2 <sup>0,00*</sup>	33 ± 11,3
Traducción	2,7	0,0 <sup>0,64</sup>	-
Tráfico subcelular y transporte	8,7	0,0 <sup>0,02*</sup>	-
Transcripción, <i>splicing</i>	6,1	0,0 <sup>0,07</sup>	-
Desconocida	38,4	28,6 <sup>0,17</sup>	9,2 ± 4,1

1: Datos de Berardini y colaboradores (2004);

2: Valor p para la probabilidad del test exacto de Fisher comparado con el genoma total

\*, valor p < 0,05.

Cinco genes seleccionados codifican proteínas involucradas en el metabolismo de los carbohidratos (10,2 %), un porcentaje significativamente mayor que el del total del genoma (2,4 %). Esta categoría incluye un gen que codifica una xiloglucano:xiloglucosil transferasa (*At3g48580*), un enzima



(E.C.2.4.1.207) que interviene en la síntesis de pared celular. También incluye una sacarosa sintasa (*At5g49190*). La sacarosa es una importante señal de regulación del proceso de desarrollo de la semilla que controla la expresión de genes de reserva (Borisjuk *et al.*, 2004).

Cinco genes relacionados con la modificación, localización o degradación de proteínas fueron seleccionados (10,2 %), dos son proteinasas (*At3g54940* y *At5g09640*). No se seleccionaron genes relacionados con la traducción a pesar de que representan el 2,7 % del total del genoma. Tampoco se encontraron genes relacionados con el transporte y el movimiento subcelular aunque representan el 8,7 % de los genes de Arabidopsis.

Se seleccionaron cuatro genes relacionados con el desarrollo (8,0 %). Dos están relacionados con la síntesis o modificación de la pared celular: *At5g59170*, que codifica una extensina, y *At3g60730*, que codifica una pectinesterasa. La presencia de genes relacionados con desarrollo no es significativamente mayor a lo observado en el total del genoma (6,0 %). Un tercer gen cuya expresión responde a auxina (*At1g48660*) también podría estar relacionado con esta categoría, pero fue incluido en el grupo de función desconocida.

Dos genes codifican posibles factores de transcripción (4,1 %): *abi3* y un gen que codifica una proteína de dedo de zinc de tipo CCHH (*At1g03790*). Otros dos genes están relacionados con energía (4,1 %).

#### 1.5 Patrones de expresión de genes durante el desarrollo de semilla y silicua

Se investigó el patrón de expresión durante el desarrollo de la semilla para cada uno de los genes seleccionados. Los datos de expresión fueron obtenidos a partir de la herramienta northern digital en Genevestigator<sup>®</sup>, utilizando los datos obtenidos de la hibridación de micromatrices con sondas provenientes de silicuas y semillas a diferentes estados de desarrollo, desde estadio globular intermedio a embriones con cotiledones verdes (Schmid *et al.*, 2005). Se utilizó el análisis SOTA (en el programa de análisis TMEV 3.1 para agrupar los patrones de expresión similares (Figura 17).

De este análisis, se pudieron distinguir cuatro grandes grupos de patrones de expresión (Figura 17):

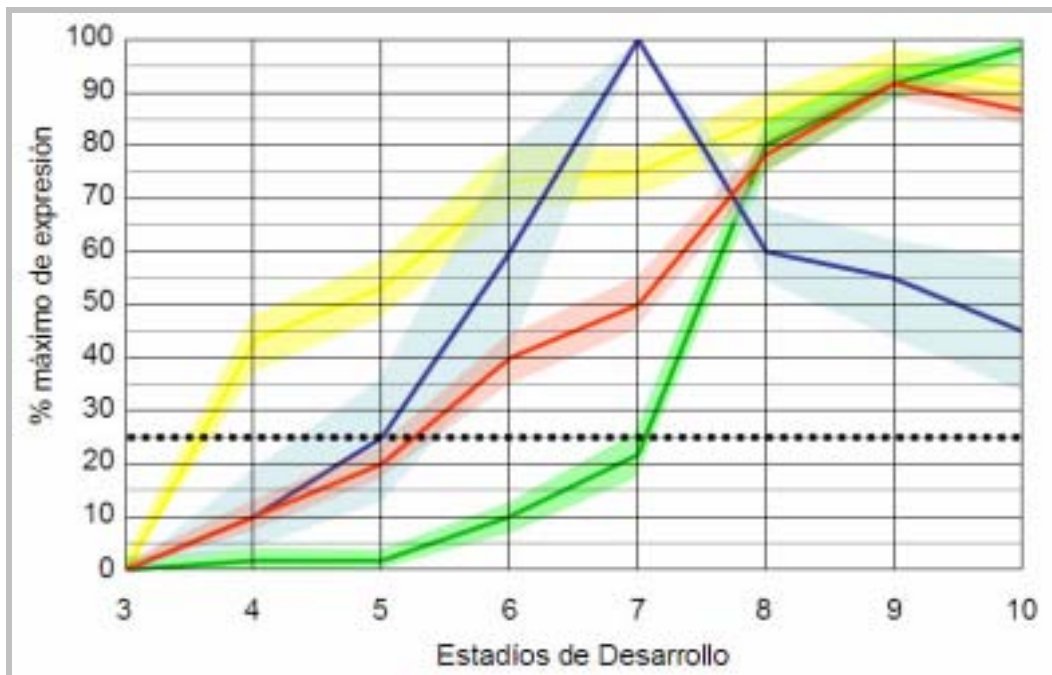


Figura 17.- Patrones de expresión durante el desarrollo de la semilla de los genes identificados mediante la sustracción *in silico*. Azul, patrón I; amarilla, patrón IIa; roja, patrón IIb; verde, patrón IIc. Líneas corresponden a los valores promedios de la expresión y las áreas oscurecidas en torno a ellas, a los errores estándar de cada una. Desarrollo de silicuas con embriones en estadios: 3, globular intermedio a corazón temprano; 4, corazón temprano a tardío; 5, corazón tardío a torpedo intermedio; 6, torpedo tardío; 7, torpedo tardío a *walking stick* temprano; 8, *walking stick* a cotiledón curvado temprano. 9, corresponde a semillas con embriones de cotiledón curvado a cotiledón verde temprano; 10, semillas con embriones con cotiledones verdes. Línea punteada representa al 25% del máximo de expresión.

- Grupo I: alta expresión en estadio temprano del desarrollo de la semilla. Los genes que alcanzaron un máximo nivel de expresión entre los estados de torpedo tardío y *walking-stick* del embrión. Este grupo incluye cinco genes: *At5g09640*, que codifica una serina carboxipeptidasa; *At5g49190*, que codifica una sacarosa sintasa; *At2g34700*, que codifica una glicoproteína rica en prolina y dos genes que codifican proteínas del tipo germina (*At3g04170* y *At3g04190*).
- Grupo II: la expresión se incrementa progresivamente alcanzando el máximo nivel después del estadio de cotiledón verde temprano. El análisis SOTA divide a este grupo en tres categorías que pueden ser

diferenciadas por el estadio al cual su nivel de expresión supera el 25% del máximo.

- IIa: Expresión muy temprana. La expresión supera el 25% del máximo antes del estadio de embrión temprano. Cuatro genes están incluidos en este grupo. *At5g48100*, que codifica una lacasa; *At4g36700*, que codifica una proteína tipo globulina; *At4g37050*, que codifica una proteína tipo patatina y *At1g62060*, que codifican una proteína de función desconocida.
- IIb: Expresión temprana. La expresión supera el 25% del máximo entre los estadios de corazón temprano y torpedo tardío. Este grupo tiene 23 genes e incluye la mayoría de los genes de “reserva de nutrientes”.
- IIc: Expresión intermedia. La expresión supera el 25% del máximo después del estado de torpedo tardío. Incluye 17 genes de diversas funciones.

#### I.6 Redundancia genética y fenotipos mutantes

Basándose en la información disponible en *The Arabidopsis Information Resource* (TAIR, [www.arabidopsis.org](http://www.arabidopsis.org)), para solo dos de los 49 genes seleccionados (4,1 %) se han descrito fenotipos mutantes (Cuadro 5). En uno de ellos (*abi3*) la mutación produce importantes alteraciones en el desarrollo del embrión y provoca, entre otras cosas, una insensibilización a la hormona ácido abscísico (Parcy *et al.*, 1994). La mutación en el otro gen (*At5g09640*) produce una anormal acumulación de sinaptoglucosa, un fenilpropanoide presente en Brasicáceas (Shirley *et al.*, 2001).

La redundancia de genes podría explicar el reducido número de mutantes entre los genes seleccionados. Muchos genes de *Arabidopsis* están en grupos en tándem o en segmentos que están duplicados (Haberer *et al.*, 2004). Se analizó cuantos de los genes de la selección eran parte una agrupación de genes en tándem o de segmentos duplicados en el genoma (Cuadro 5). 11 de los genes seleccionados (22%) están duplicados, lo que es levemente superior que lo observado en el genoma completo (17%) (diferencia no significativa para  $\alpha = 0.05$ ).

**Capítulo II. Genes que codifican proteínas con repeticiones anquirina y dominios transmembrana implicados en la embriogénesis de *Arabidopsis thaliana*.**

Los recientes progresos en la secuenciación de genomas completos han conducido a un enriquecimiento rápido de las bases de secuencias con una variedad sin precedentes de tipos de proteínas deducidas para la mayoría de las cuales se desconoce su función (Tatusov *et al.*, 2000). A partir de estos datos se ha observado que, a pesar del número prácticamente ilimitado de secuencias posibles de proteínas, el número de formas básicas existentes parece no solamente ser finito, sino relativamente pequeño, con probablemente no mucho más de 10.000 formas básicas (Koonin *et al.*, 2002). Por otra parte, la distribución del número de genes que codifican estos tipos básicos de proteínas es poco homogénea, habiendo unos pocos tipos muy abundantes y la mayoría muy escasos.

Las formas básicas de las proteínas quedan definidas en muchas ocasiones por lo que se conoce como dominios proteícos (Vogel *et al.*, 2004). Los dominios de proteína son unidades elementales de la estructura y evolución de las proteínas. No hay acuerdo unánime en cuanto a la definición de dominio de proteína. Una posible definición (Przytycka *et al.*, 2005) es que los dominios son unidades evolutivas conservadas que: (1) se pliegan de manera independiente, (2) aparecen en diferentes proteínas acompañadas de otros dominios diversos, y (3) son unidades mínimas que satisfacen (1) y (2). Los distintos dominios pueden combinarse entre sí, formando lo que se conoce como proteínas multidominio. Cerca de dos tercios de las proteínas en Procariotas y el ochenta por ciento en Eucariotas son proteínas multidominio (Apic *et al.*, 2001). Como promedio, las proteínas tienen de dos a tres dominios, pero hay proteínas para las cuales se han identificado más de cien (Gerstein, 1998; Teichmann, 1998). Las proteínas multidominio tienen características únicas que las dotan de una importante significación evolutiva. En ellas pueden coexistir una gran cantidad de funciones haciéndolas componentes imprescindibles de redes reguladoras o estructurales donde son esenciales las interacciones múltiples (proteína-proteína, proteína-ligando, proteína-DNA, etc.) (Tordai *et al.*, 2005). La formación de proteínas multidominio contribuye de manera esencial al incremento de la complejidad de los organismos (Patthy, 2003).

El objetivo de esta parte del trabajo es el estudio en *Arabidopsis*, y en general en las plantas, de uno de los dominios proteicos más abundantes, las repeticiones anquirina (ANK). Las repeticiones ANK son dominios que presentan una estructura básica de 33 aminoácidos repetida en tándem al menos dos veces (Bork, 1993). Se han identificado repeticiones ANK en numerosas proteínas involucradas en muy diversas funciones (ver Introducción general) y se cree que funcionan como lugares de unión proteína-proteína. Primero, se ha determinado la secuencia consenso, segundo se han catalogado las familias de proteínas que poseen repeticiones ANK, y, por último, se ha estudiado más a fondo una de ellas, las proteínas multidominio con repeticiones anquirina y dominios transmembrana, tanto a nivel genético como de patrón de expresión.

## II.1 Repeticiones anquirina en *Arabidopsis*

Mediante búsquedas en las bases de secuencias se identificaron todos los genes de *Arabidopsis* que codifican proteínas con repeticiones ANK. Para ello se empleó un sistema iterativo. La búsqueda inicial se basó en la secuencia ANK consenso determinada para proteínas animales (Bork, 1993) que utilizaban los programas de búsqueda de dominios SMART v3.5 (<http://smart.embl-heidelberg.de/>), TAIR *Protein Search* (<http://www.arabidopsis.org>). Estos programas nos proporcionaron una colección inicial de proteínas de *Arabidopsis* con repeticiones ANK. Una vez obtenida esta primera colección de secuencias, se identificaron todas las posibles repeticiones ANK mediante el programa REP v1.1 (<http://www.embl-heidelberg.de/~andrade/papers/rep/search.html>; Andrade *et al.*, 2000), que permite el uso de parámetros de búsqueda menos restrictivos.

En un segundo paso, cada una de las repeticiones ANK identificadas fue comparada con las bases de secuencias mediante el programa TBLASTN (<http://www.ncbi.nlm.nih.gov/BLAST/>), se añadieron las nuevas proteínas a la selección y se analizaron de nuevo con REP v1.1. Se repitieron los mismos pasos con las nuevas secuencias identificadas, tantas veces hasta que no fue posible identificar nuevas repeticiones ANK, cosa que se produjo a la tercera iteración.

Mediante este sistema se aseguró el aislamiento de la totalidad de las posibles repeticiones ANK codificadas en el genoma de Arabidopsis. Sin embargo, también se corría el riesgo de incluir secuencias parecidas pero erróneas. Con el fin de eliminar todas aquellas secuencias que no correspondían a repeticiones ANK reales solamente se tuvieron en cuenta las repeticiones ANK presentes en grupos de al menos 2 repeticiones separadas por no más de 20 aminoácidos, dado que la capacidad funcional de las repeticiones ANK depende de la existencia de al menos dos repeticiones consecutivas. Por lo tanto, se eliminaron todas aquellas repeticiones aisladas y también se eliminaron las repeticiones parciales y las posibles repeticiones que no conservaran al menos dos de los seis aminoácidos más conservados de la secuencia consenso definida para Arabidopsis (II.1.1).

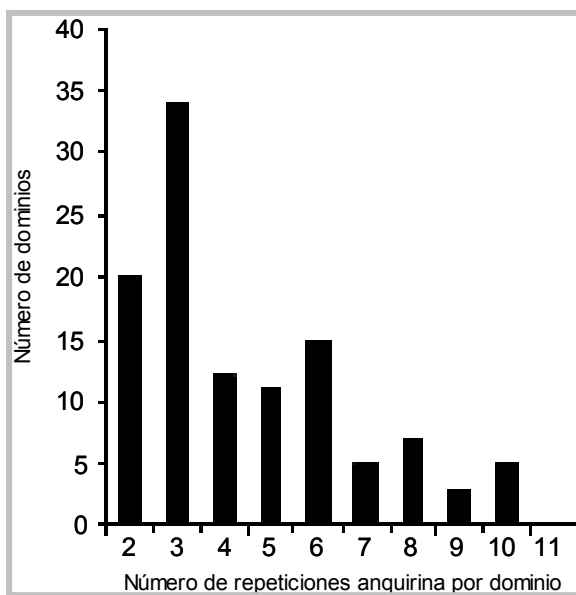


Figura 18.- Gráfico de frecuencia del número de repeticiones anquirina por agrupación en tándem.

Finalmente, estos análisis permitieron identificar un total de 509 repeticiones ANK codificadas por 105 genes. El número de repeticiones ANK en tándem oscila entre 2 y 10, siendo el promedio de 4.5 (Figura 18). Algunas de las proteínas contienen 2 grupos de repeticiones ANK separados.

### II.1.1 Secuencia consenso de las repeticiones anquirina en Arabidopsis

Las 509 repeticiones ANK identificadas fueron alineadas mediante CLUSTALW (<http://www.ebi.ac.uk/clustalw/>) y se determinó su secuencia consenso (Cuadro 6).

Cuadro 6.- Secuencia consenso de las repeticiones anquirina de *Arabidopsis thaliana*.

Posición del aminoácido	Animales		<i>Arabidopsis thaliana</i>		
	Bork (1993)	Michaely y Bennett, (1992)	Consenso	Aminoácidos más abundantes	Porcentaje (%)
1	-	-	Hidrofílico	D	21
2	<i>t</i>	G	G	G	63
3	-	-	Hidrofílico	-	-
4	T ó S	T	T	T	56
5	P	A, P	P, A	P ; A	37 ; 29
6	L	L	L	L	77
7	H	H	H	H	64
8	<i>h</i>	L, I, V	Hidrofóbico	L	23
9	A	A	A	A	84
10	<i>h</i>	A, S	A	A ; V	48 ; 20
11	-	R, Q, K	Hidrofílico	-	-
12	-	-	Hidrofílico	-	-
13	<i>t</i>	G, N	G	G	62
14	<i>t</i>	H, N	Hidrofílico	H	24
15	-	V, L, T	-	-	-
16	<i>t</i>	E, D	Hidrofílico	E	29
17	<i>h</i>	V, I, M	Hidrofóbico	V ; I	23 ; 20
18	<i>h</i>	V, A	V	V	45
19	<i>t</i>	K, E, R	Hidrofílico	K	27
20	-	L, V	-	L	18
21	L	L	L	L	63
22	L	L	L	L	46
23	<i>t</i>	D, K, Q, E	Hidrofílico	E	19
24	-	-	Hidrofílico	-	-
25	<i>t</i>	G	-	G	30
26	-	A	-	A ; P	24 ; 18
27	-	D, N, S	Hidrofílico	D	20
28	-	V, P, I	-	L	19
29	-	N, D	-	-	-
30	-	A	-	-	-
31	-	-	-	-	-
32	-	T, D, N	Hidrofílico	D	19
33	-	K	Hidrofílico	N	19

La secuencia consenso de *Arabidopsis thaliana* contiene sólo un aminoácido si éste representa más del 40% y una clase de aminoácido, si representa más del 60%. Muchos aminoácidos abundantes sólo son indicados si representan el 18% o más del total. Los porcentajes están calculados en base a las 509 repeticiones anquirina identificadas en este estudio. -, señala que no hay ningún aminoácido o tipo de aminoácido especial en esa posición. *t*, representa a un aminoácido polar o que tiende a formar giros. *h*, indica un aminoácido hidrofóbico. Las letras mayúsculas corresponden a aminoácidos conservados de acuerdo al código de una letra.

Ninguna de las posiciones estaba completamente conservada, pero se identificaron siete posiciones cuyos residuos poseían al menos un 50% de conservación. Cinco de ellos están localizados entre las posiciones 2 y 9, que es la región más conservada. Estos siete residuos ya habían sido previamente señalados como muy conservados al determinar la secuencia consenso de animales (Michaely y Bennett, 1992; Bork, 1993). También se ha observado, tanto en *Arabidopsis* como en animales, una estricta conservación de posiciones hidrofílicas e hidrofóbicas (Michaely y Bennett, 1992; Bork, 1993).



## II.1.2 Proteínas que contienen repeticiones anquirina en Arabidopsis.

Como se ha comentado anteriormente, se identificaron 105 genes en el genoma de Arabidopsis que codifican proteínas con repeticiones ANK (Cuadro 7). Para mayor comodidad, a cada gen le fue asignado un número de entrada (EN).

Cuadro 7.- Genes de *Arabidopsis thaliana* que codifican proteínas que contienen repeticiones anquirina

EN	Número Atg	Grupo y Familia <sup>(a)</sup>	Cr <sup>(b)</sup>	Posición en el mapa cromosómico (pb)	Nombre	Función	Referencia
<i>Grupo A: Proteínas con dominios Anquirina-Transmembrana</i>							
1	At1g03670	A1	1	914221-916221 c	n.d.	n.d.	n.d.
2	At4g03440	A1	4	1524197-1527131 c	n.d.	n.d.	n.d.
3	At4g03450	A1	4	1529444-1531734 c	n.d.	n.d.	n.d.
4	At4g03460	A1	4	1536402-1540109 c	n.d.	n.d.	n.d.
5	At4g03470	A1	4	1542057-1544281 c	n.d.	n.d.	n.d.
6	At4g03480	A1	4	1546022-1548869 c	n.d.	n.d.	n.d.
7	At4g03490	A1	4	1549616-1552782 c	n.d.	n.d.	n.d.
8	At4g03500	A1	4	1553090-1556569	n.d.	n.d.	n.d.
9	At4g05040	A1	4	2578687-2581795	n.d.	n.d.	n.d.
10	At4g14390	A1	4	8289640-8292079	n.d.	n.d.	n.d.
11	At4g14400	A1	4	8294448-8298598	ACD6	Respuesta a patógenos	Lu <i>et al.</i> , 2003
12	At1g14480	A2	1	4956399-4957931	n.d.	n.d.	n.d.
13	At1g14500	A2	1	4960370-4961775	n.d.	n.d.	n.d.
14	At4g10720	A2	4	6607875-6609354	n.d.	n.d.	n.d.
15	At4g11000	A2	4	6731016-6732460	n.d.	n.d.	n.d.
16	At5g15500	A2	5	5031679-5033503 c	n.d.	n.d.	n.d.
17	At5g51160	A2	5	20809506-20811000	n.d.	n.d.	n.d.
18	At5g54610	A2	5	22202005-22203733 c	n.d.	n.d.	n.d.
19	At5g54620	A2	5	22204987-22206972 c	n.d.	n.d.	n.d.
20	At1g10340	A3	1	3390477-3392483 c	n.d.	n.d.	n.d.
21	At1g34050	A3	1	12393474-12395985	n.d.	n.d.	n.d.
22	At2g24600	A3	2	10459238-10461603 c	n.d.	n.d.	n.d.
23	At5g50140	A3	5	20413082-20415423	n.d.	n.d.	n.d.
24	At5g54700	A3	5	22240322-22242735 c	n.d.	n.d.	n.d.
25	At5g54710	A3	5	22244800-22247726 c	n.d.	n.d.	n.d.
26	At5g54720	A3 <sup>(c)</sup>	5	22249521-22250078	n.d.	n.d.	n.d.
27	At1g05640	A4	1	1687435-1689500 c	n.d.	n.d.	n.d.
28	At1g07710	A4	1	2386272-2387983 c	n.d.	n.d.	n.d.
29	At2g01680	A4	2	306384-308629	n.d.	n.d.	n.d.
30	At2g31820	A4	2	13537427-13539639	n.d.	n.d.	n.d.
31	At3g09550	A4	3	2932468-2934359	n.d.	n.d.	n.d.
32	At3g12360	A4	3	3934085-3936701	n.d.	n.d.	n.d.
33	At5g02620	A4	5	589536-591675	n.d.	n.d.	n.d.
34	At5g60070	A4	5	24207666-24209796 c	n.d.	n.d.	n.d.
35	At3g18670	A5	3	6424141-6426477 c	n.d.	n.d.	n.d.
36	At3g54070	A5	3	20032308-20034581 c	n.d.	n.d.	n.d.
37	At5g04690	A5	5	1349644-1352526 c	n.d.	n.d.	n.d.
38	At5g35830	A5 <sup>(c)</sup>	5	14017543-14018646 c	n.d.	n.d.	n.d.
39	At2g14250	A6 <sup>(c)</sup>	2	6044056-6044791	n.d.	n.d.	n.d.
40	At5g20350	A6	5	6876591-6881272	n.d.	n.d.	n.d.
<i>Grupo B: Proteínas con sólo dominios anquirina</i>							
41	At1g04780	B1	1	1340890-1342964 c	n.d.	n.d.	n.d.
42	At1g11740	B1	1	3963277-3966824 c	n.d.	n.d.	n.d.
43	At1g62050	B1	1	23004028-23004401 c	n.d.	n.d.	n.d.
44	At3g04470	B1	3	1189646-1191191 c	n.d.	n.d.	n.d.
45	At3g24210	B1	3	8753481-8756072 c	n.d.	n.d.	n.d.
46	At2g17390	B2	2	7562952-7565120	ART2	Respuesta a patógenos	Peck <i>et al.</i> , 2001
47	At4g35450	B2	4	16839559-16842082	ARP2	Respuesta a patógenos	Yan <i>et al.</i> , 2002
48	At5g40160	B3	5	16079714-16081558 c	EMB506	Embriogénesis	Albert <i>et al.</i> , 1999
49	At5g66055	B3	5	26434382-26436490 c	AKR	Diferenciación celular regulada por luz	Zhang <i>et al.</i> , 1992
50	At5g07840	B4	5	2506657-2508374 c	n.d.	n.d.	n.d.
51	At5g61230	B4	5	24644418-24646174	n.d.	n.d.	n.d.

Cuadro 7.- Genes de *Arabidopsis thaliana* que docifican proteínas que contienen repeticiones anquirina (continuación)

EN	Número Atg	Grupo y Familia <sup>(*)</sup>	Cr <sup>(b)</sup>	Posición en el mapa cromosómico (pb)	Nombre	Función	Referencia
<i>Grupo B: Proteínas con sólo dominios anquirina (continuación)</i>							
52	At3g01750	B5	3	270236-272873	n.d.	n.d.	n.d.
53	At3g04140	B5	3	1086998-1089236	n.d.	n.d.	n.d.
54	At5g65860	B6	5	26364932-26365912 c	n.d.	n.d.	n.d.
55	At4g19150	B7	4	10471346-10472753 c	n.d.	n.d.	n.d.
56	At2g03430	B8	2	1036029-1037613 c	n.d.	n.d.	n.d.
57	At5g12320	B9	5	3982696-3984049	n.d.	n.d.	n.d.
58	At3g09890	B10	3	3032607-3034376	n.d.	n.d.	n.d.
<i>Grupo C: Proteínas con dominio BTB</i>							
59	At1g64280	C1	1	23918273-23921091 c	NPR1	<i>non-expresser of PR genes</i>	Cao <i>et al.</i> , 1997, 1998
60	At2g41370	C1	2	17244823-17247518 c	NPR1	<i>non-expresser of PR genes</i>	Cao <i>et al.</i> , 1997, 1998
61	At3g57130	C1	3	21158814-21161006	NPR1	<i>non-expresser of PR genes</i>	Cao <i>et al.</i> , 1997, 1998
62	At4g19660	C1	4	10696276-10698253 c	NPR1	<i>non-expresser of PR genes</i>	Cao <i>et al.</i> , 1997, 1998
63	At4g26120	C1	4	13236457-13238496	NPR1	<i>non-expresser of PR genes</i>	Cao <i>et al.</i> , 1997, 1998
64	At5g45110	C1	5	18246137-18248749	NPR1	<i>non-expresser of PR genes</i>	Cao <i>et al.</i> , 1997, 1998
65	At2g04740	C2	2	1657134-1659560	n.d.	n.d.	n.d.
<i>Grupo D: Quinasas</i>							
66	At1g14000	D1	1	4797355-4800278	APK	Proteína quinasa	Chinchilla <i>et al.</i> , 2003
67	At2g31800	D1	2	13527860-13530723 c	APK	Proteína quinasa	Chinchilla <i>et al.</i> , 2003
68	At3g58760	D1	3	21739733-21742905	APK	Proteína quinasa	Chinchilla <i>et al.</i> , 2003
69	At3g59830	D1	3	22114075-22116300 c	APK	Proteína quinasa	Chinchilla <i>et al.</i> , 2003
70	At4g18950	D1	4	10375392-10378400	APK	Proteína quinasa	Chinchilla <i>et al.</i> , 2003
71	At2g43850	D2	2	18166259-18169061	n.d.	n.d.	n.d.
72	At5g13530	D3	5	4345621-4351150	n.d.	n.d.	n.d.
<i>Grupo E: Proteínas con dedos de zinc</i>							
73	At2g40140	E1	2	16779238-16781735	n.d.	n.d.	n.d.
74	At2g41900	E1	2	17497490-17501000	n.d.	n.d.	n.d.
75	At3g55980	E1	3	20787274-20789821	n.d.	n.d.	n.d.
76	At5g12850	E1	5	4056194-4059583	n.d.	n.d.	n.d.
77	At5g58620	E1	5	23710566-23713448	n.d.	n.d.	n.d.
78	At3g28880	E2	3	10893672-10897009	n.d.	n.d.	n.d.
<i>Grupo F: Proteínas de canales de potasio</i>							
79	At2g25600	F1	2	10901681-10905447	AKT	<i>Voltage-dependent K<sup>+</sup> channels</i>	Pilot <i>et al.</i> , 2003
80	At2g26650	F1	2	11338365-11343579 c	AKT	<i>Voltage-dependent K<sup>+</sup> channels</i>	Pilot <i>et al.</i> , 2003
81	At3g02850	F1	3	619171-623565 c	AKT	<i>Voltage-dependent K<sup>+</sup> channels</i>	Pilot <i>et al.</i> , 2003
82	At4g22200	F1	4	11746525-11750710 c	AKT	<i>Voltage-dependent K<sup>+</sup> channels</i>	Pilot <i>et al.</i> , 2003
83	At4g32500	F1	4	15681128-15685220	AKT	<i>Voltage-dependent K<sup>+</sup> channels</i>	Pilot <i>et al.</i> , 2003
84	At5g37500	F1	5	14906923-14912134 c	AKT	<i>Voltage-dependent K<sup>+</sup> channels</i>	Pilot <i>et al.</i> , 2003
<i>Grupo G: Proteínas Ring Finger</i>							
85	At3g23280	G1	3	8321397-8324435	n.d.	n.d.	n.d.
86	At4g14365	G1	4	8271460-8273733 c	n.d.	n.d.	n.d.
87	At5g07270	G2	5	2280822-2283594	n.d.	n.d.	n.d.
88	At5g57740	G2	5	23411811-23414876 c	n.d.	n.d.	n.d.
89	At2g28840	G3	2	12385436-12387819	n.d.	n.d.	n.d.
<i>Grupo H: Proteínas con dominios de activación ARF GTPasa</i>							
90	At1g10870	H1	1	3616905-3623612	n.d.	n.d.	n.d.
91	At1g60860	H1	1	22466771-22473166 c	n.d.	n.d.	n.d.
92	At5g13300	H1	5	4255605-4262317	n.d.	n.d.	n.d.
93	At5g61980	H1	5	24911698-24916404	n.d.	n.d.	n.d.
<i>Grupo I: Proteínas con motivos de unión a calmodulina</i>							
94	At1g67310	I1	1	25263707-25268651 c	CAMTA	<i>calmodulin-binding transcr. activators</i>	Bouché <i>et al.</i> , 2002
95	At2g22300	I1	2	9478472-9484147	CAMTA	<i>calmodulin-binding transcr. activators</i>	Bouché <i>et al.</i> , 2002
96	At5g09410	I1	5	2921458-2927417	CAMTA	<i>calmodulin-binding transcr. activators</i>	Bouché <i>et al.</i> , 2002
97	At5g64220	I1	5	25703472-25709441	CAMTA	<i>calmodulin-binding transcr. activators</i>	Bouché <i>et al.</i> , 2002
<i>Grupo J: Proteínas de unión a Acil-CoA</i>							
98	At4g27780	J1	4	13847555-13849893	ACBT	Proteína de unión a Acil-CoA	Chye <i>et al.</i> , 2000
99	At5g53470	J1	5	21727577-21729836	ACBT	Proteína de unión a Acil-CoA	Chye <i>et al.</i> , 2000
<i>Grupo K: Proteína con cromodominio</i>							
100	At2g47450	K1	2	19479751-19481384	CAO	Proteína de unión a cloroplasto	Klimyuk <i>et al.</i> , 1999

Cuadro 7.- Genes de *Arabidopsis thaliana* que docifican proteínas que contienen repeticiones anquirina (continuación).

EN	Número Atg	Grupo y Familia <sup>(a)</sup>	Cr <sup>(b)</sup>	Posición en el mapa cromosómico (pb)	Nombre	Función	Referencia
<i>Grupo L: Helicasa</i>							
101	At1g06670	L1	1	2040432-2047611	DEAH	Helicasa DNA/RNA	Isono <i>et al.</i> , 1999
<i>Grupo M: Proteína con dominio RCC-1</i>							
102	At3g03790	M1	3	962004-968156	n.d.	n.d.	n.d.
<i>Grupo N: Proteína con repeticiones tetratricopéptidos</i>							
103	At3g04710	N1	3	1278085-1281124	n.d.	n.d.	n.d.
<i>Grupo O: Proteína con dominio PH</i>							
104	At5g14230	O1	5	4593810-4595967	n.d.	n.d.	n.d.
<i>Grupo P: Proteína con motivo asociado a ATPasa</i>							
105	At3g24530	P1	3	8945476-8947933 c	n.d.	n.d.	n.d.

<sup>(a)</sup>: dentro de cada grupo (letra mayúscula), las familias (número) se basan en similitud de secuencia.

<sup>(b)</sup>: Cr, cromosoma.

<sup>(c)</sup>: proteínas que sólo tienen repeticiones anquirina, pero son agrupadas con proteínas que contienen dominios adicionales por similitud de secuencias.

pb: pares de bases; c: cadena complementaria; n.d.: no determinado.

Las proteínas codificadas por estos 105 genes fueron analizadas mediante el programa SMART v3.5 (<http://smart.embl-heidelberg.de/>) para determinar la presencia de otros dominios proteicos. Atendiendo a la presencia de dominios proteicos se identificaron 16 grupos (Cuadro 8). Utilizando alineamientos múltiples de secuencias (CLUSTALW), se identificaron similitudes de secuencia y cada uno de estos grupos fueron subdivididos en familias.

Se encontraron 37 genes que codifican proteínas que poseen repeticiones ANK y dominios transmembrana (Grupo A) y 21 que codifican proteínas sólo con repeticiones ANK (grupo B). Sin embargo, los genes EN26, EN38 y EN39, que pertenecen estructuralmente al grupo B, fueron incluidos en el grupo A debido a que sus secuencias son mucho más similares a la región de repeticiones ANK de algunos genes del grupo A que a las del resto del grupo B. Probablemente corresponden a formas truncadas de proteínas del grupo A.

Del total de proteínas clasificadas (105), sólo 31 han sido previamente caracterizadas. Los grupos que poseen mayor número de genes son aquellos que a su vez tienen una más baja proporción de genes estudiados. Sólo uno de los cuarenta genes en el grupo A ha sido estudiado (Lu *et al.*, 2003). Por esta razón, se determinó estudiar este grupo en mayor profundidad (Apartado II.2).

Cuadro 8.- Tipos de proteínas que contienen repeticiones anquirina en *Arabidopsis thaliana*.

Grupo	Descripción	Número de genes	Representación esquemática
A	Proteínas con dominios transmembrana	40	
B	Proteínas con solo repeticiones anquirina	18	
C	Proteínas con repeticiones anquirina y dominios BTB	7	
D	Quinasas	7	
E	Proteínas con dedos de zinc	6	
F	Proteínas de canales de potasio	6	
G	Proteínas con Ring Finger	5	
H	Proteínas con dominios de activación ARF GTPasa	4	
I	Proteínas con motivo de unión a calmodulina	4	
J	Proteínas con dominio de unión a Acil CoA	2	
K	Proteína con cromodominio	1	
L	Helicasa	1	
M	Regulador de condensación del cromosoma	1	
N	Proteína que contiene repeticiones de Tetratricopéptidos	1	
O	Proteína con dominio PH	1	
P	Proteína con dominio asociado a ATPasa	1	

**A** Repetición anquirina; **T** Dominio transmembrana; **BTB** *Broad-Complex, Tramtrack and Bric a Brac*;  
**KIN** Dominio de proteína quinasa; **Z** Dominio dedo de Zinc; **PIR** Dominio proteico de transporte de iones;  
**cNMP** Dominio de unión de nucleótido monofosfato cíclico; **R** *Ring finger*; **BAR** Dominio BAR ;  
**PH** Dominio de homología Pleckstrin; **ARF** Dominio de activación ARF GTPasa; **CG-1** Dominio CG-1;  
**Iq** Motivo de unión a calmodulina; **ACBP** Dominio de proteína de unión a Acil CoA; **C** Cromodominio;  
**N** Dominio de unión a ácidos nucleicos ; **DEXD** Superfamilia de helicasa tipo DEAD; **HE** Dominio C-terminal de la superfamilia helicasa; **HA** Dominio asociado a helicasa;  
**D** Dominio regulador de la condensación del cromosoma; **Tt** Repeticiones tetratricopéptido; **AAA** Dominio asociado a ATPasa

### II.1.2.1 Proteínas con sólo repeticiones anquirina

21 genes de *Arabidopsis* codifican proteínas cuyo único motivo reconocible son repeticiones ANK (Cuadro 8): Familias B1 a B10 y genes EN26, EN 38 y EN39. El tamaño de las proteínas codificadas va desde 144 a 664 aminoácidos y el número de repeticiones ANK de 2 a 10. En el caso más extremo, las repeticiones ANK comprenden el 87% de la proteína (EN57), en otros casos los grupos de repeticiones ANK están concentrados en las regiones N- o C-terminal. Cuatro de estos genes han sido caracterizados previamente. Dos de ellos (EN46 y EN47) codifican proteínas similares a ANK1 de tabaco y participan en respuesta de defensa contra patógenos (Peck *et al.*, 2001; Yan *et al.*, 2002), mientras que las funciones de EN48 y EN49 están relacionados con embriogénesis y desarrollo (Zhang *et al.*, 1992; Albert *et al.*, 1999).

### II.1.2.2 Grupo con dominio BTB

Siete genes codifican proteínas con dominios BTB y repeticiones ANK. El dominio BTB, también denominado POZ (poxvirus y *zinc finger*), es conocido por ser un motivo de interacción proteína-proteína encontrado en la región N-terminal de diversos factores de transcripción del tipo C2H2, así como en canales de potasio del tipo Shaw (Bardwell y Treisman, 1994). Este grupo de genes está dividido en dos familias (C1 y C2) con diferente orden en los dominios. La familia C1 contiene 6 proteínas en las que el dominio BTB está localizado en el N-terminal y las repeticiones ANK en el C-terminal. Estos genes codifican proteínas similares a NPR1 que participa en el control de la respuesta de resistencia sistémica adquirida a un amplio espectro de patógenos (Cao *et al.*, 1997). Se han encontrado proteínas similares a NPR1 en otras especies vegetales tales como arroz, *Brassica* y tabaco, pero no se ha encontrado ninguna proteína con similar estructura de dominios en animales u hongos. La familia C2 por su parte, contiene un gen que codifica una proteína con los dominios en el orden inverso a C1. Se han encontrado proteínas con una similar organización de dominios en animales y hongos, pero sus funciones son desconocidas excepto para una proteína de unión al factor de elongación 1A de humanos (Unoki y Nakamura, 2001).

### II.1.2.3 Proteína quinasas

Los genes que codifican proteínas con repeticiones ANK y dominios quinasa son siete y se dividen en tres familias. En las familias D1 (5 genes) y D2 (1 gen), las repeticiones ANK están localizadas en la mitad N-terminal de la proteína y el dominio quinasa en el C-terminal. Las proteínas D1 son similares a la proteína anquirina-quinasa (*APK*) de *Medicago* (Chinchilla *et al.*, 2003). También existen homólogos estructurales de D1 y D2 en animales, por ejemplo, la quinasa cardiaca humana con repeticiones ANK (Acc. NM\_015978). La familia D3 contiene un gen (EN72) que codifica una proteína con un dominio *ring finger* en el N-terminal, un dominio quinasa en el centro y repeticiones ANK en C-terminal. El gen EN66 de la familia D1 parece codificar una tirosina-quinasa, pero la especificidad quinasa de las otras proteínas es desconocida.

### II.1.2.4 Proteínas con dedos de zinc

Seis genes codifican proteínas con repeticiones ANK y dominios de dedos de zinc. Están divididos en dos familias (E1 y E2). La familia E1 contiene 5 genes que codifican proteínas con grupos cortos de 2 ó 3 repeticiones ANK en el extremo N-terminal y 1 ó 2 dedos de zinc en la parte central de la proteína. Hay genes de arroz que codifican proteínas similares. La familia E2 contiene un gen que codifica una proteína con dominios similares pero tiene un conjunto de seis repeticiones ANK. No se ha determinado ninguna función para estas proteínas.

### II.1.2.5 Canales de Potasio

La primera proteína vegetal descrita que contenía repeticiones ANK fue AKT1 (EN80), que codifica una proteína similar a un canal de K<sup>+</sup> tipo *shaker* localizado en la membrana plasmática (Sentenac *et al.*, 1992). Los canales de potasio *shaker* juegan un importante papel en la absorción de K<sup>+</sup> del suelo. El genoma de *Arabidopsis* contiene nueve genes que codifican canales de potasio *Shaker* y seis de ellos contienen repeticiones ANK (Pilot *et al.*, 2003). Los canales *Shaker* vegetales comparten una estructura común: una parte central hidrofóbica compuesta por seis fragmentos transmembrana, una larga región

citoplasmática en C-terminal que contiene un posible dominio de unión a nucleótidos, y un dominio KHA, que es rico en aminoácidos hidrofóbicos y ácidos y juegan un importante papel en el proceso de tetramerización del canal (Ehrhardt *et al.*, 1997; Daram *et al.*, 1997; Mouline, *et al.*, 2002). Muchos canales, pero no todos, también contienen repeticiones ANK entre el posible dominio de unión cíclico y el dominio KHA. Se han descrito genes que codifican proteínas similares en muchas otras especies vegetales, incluyendo dicotiledóneas y monocotiledóneas (Pilot *et al.*, 2003).

#### II.1.2.6 Proteínas *Ring finger*

Seis genes de *Arabidopsis* codifican proteínas con dominios *ring finger* y repeticiones ANK. Uno de ellos también tiene un dominio quinasa y por esta razón ha sido incluido en la familia D3 (EN72). Los cinco restantes contienen de 4 a 6 repeticiones ANK en el extremo N-terminal y un *ring finger* en C-terminal. Están divididos en tres familias (G1 – G3). Genes de arroz y de animales codifican proteínas similares, pero sus funciones son desconocidas.

#### II.1.2.7 Proteínas con dominios de activación ARF GTPasa

Se identificaron cuatro genes de *Arabidopsis* con una organización similar (Familia H1), desde N a C-terminal: un dominio BAR, un dominio PH, un dominio de activación GTPasa y dos a tres repeticiones ANK. No se conoce su función.

#### II.1.2.8 Proteínas que contienen motivos de unión a Calmodulina

Cuatro genes de *Arabidopsis* (familia I1) codifican proteínas con una organización similar: un dominio CG-1 en N-terminal, dos o tres repeticiones ANK en la región central, y dos motivos de unión a calmodulina en C-terminal. Estas proteínas se conocen como CAMTAs (*Calmodulin-binding transcription activators*) y se han descrito en *Arabidopsis* y otras especies vegetales (Bouché *et al.*, 2002). Los dominios CG-1 están muy conservados con más de 130 residuos de aminoácidos que contienen una predicción de señal de localización nuclear bipartita. Un clon parcial de cDNA de perejil codifica una proteína de secuencia específica de unión a DNA (da Costa e Silva, 1994).

#### II.1.2.9 Proteínas de unión a Acil-CoA

Se identificaron dos genes que codifican proteínas con motivos de unión a acil-CoA, repeticiones ANK en la región C-terminal y motivos transmembrana en la N-terminal (Chye *et al.*, 2000) (Familia J1).

#### II.1.2.10 Proteína con cromodominio

El gen CAO (*chlorophyll a/b binding protein harvesting-organelle specific protein*) de Arabidopsis codifica una proteína con cromodominios y repeticiones ANK (Klimyuk *et al.*, 1999). Es un gen nuclear que codifica una proteína de reconocimiento de señal de cloroplasto que es parte de un complejo de proteínas. Las repeticiones ANK son necesarias para la formación del complejo.

#### II.1.2.11 Helicasa

El gen EN101 codifica una proteína similar a la familia DEAH de las helicasas de RNA/DNA (Isono *et al.*, 1999). La proteína contiene dos repeticiones ANK.

#### II.1.2.12 Otras proteínas

Se encontraron cuatro genes más que codifican proteínas con repeticiones ANK y que también contienen algunos motivos proteicos reconocidos como, cuatro dominios RCC-1 (EN102), tres repeticiones tetratricopéptido (EN103), un dominio PH (EN104) y un motivo asociado a ATPasa (EN105). Ninguno de éstos o proteínas similares, han sido caracterizados en vegetales y sus funciones son desconocidas.



## II.2 Genes que codifican proteínas con repeticiones anquirina y dominios transmembrana en Arabidopsis

De entre los grupos de genes de Arabidopsis que codifican proteínas con repeticiones ANK, el grupo A, que corresponde a proteínas con repeticiones ANK y dominios transmembrana es el más abundante y, al mismo tiempo, uno de los grupos más desconocidos.

### II.2.1 Análisis filogenético

#### II.2.1.1 Genes que codifican proteínas ANKTM en Arabidopsis

Las proteínas deducidas del grupo A contienen de 4 a 11 repeticiones ANK en la región N-terminal y de 2 a 5 dominios transmembrana en C-terminal (Cuadro 8). A estas proteínas se les denominó proteínas ANKTM (ANKyrin TransMembrane). Curiosamente, ningún gen de Arabidopsis codifica proteínas con los mismos dominios pero en orden inverso.

El grupo A contiene 40 genes, 37 codifican proteínas con repeticiones ANK y dominios transmembrana, y tres sólo las repeticiones, aunque presentan mucha mayor similitud de secuencia con genes del grupo A que con los del B, que sería el que les correspondería por presencia de dominios reconocibles. Por tanto, deducimos que corresponden a formas truncadas de genes del grupo A.

Las secuencias de aminoácidos deducidas de las proteínas ANKTM se alinearon mediante el programa CLUSTALW y construyó un árbol por el método de *neighbor joining*. Los resultados muestran que el grupo A está dividido en 6 familias. Se obtuvieron resultados similares utilizando únicamente las secuencias de las regiones con repeticiones ANK (Figura 19.A), o las regiones con los dominios transmembrana.

Las secuencias de proteínas deducidas para cada una de las familias se alienaron por separado. Basándose en estos alineamientos se estableció un modelo de consenso de proteína para cada familia (Figura 19.B). Algunas de las repeticiones ANK están presentes en muchas o todas las proteínas de la familia, pero en otras no son completas debido a inserciones, deleciones o

mutaciones. La posición y número de dominios transmembrana está conservado en todas las proteínas de la misma familia, excepto para las tres proteínas truncadas (EN26, EN38 y EN39).

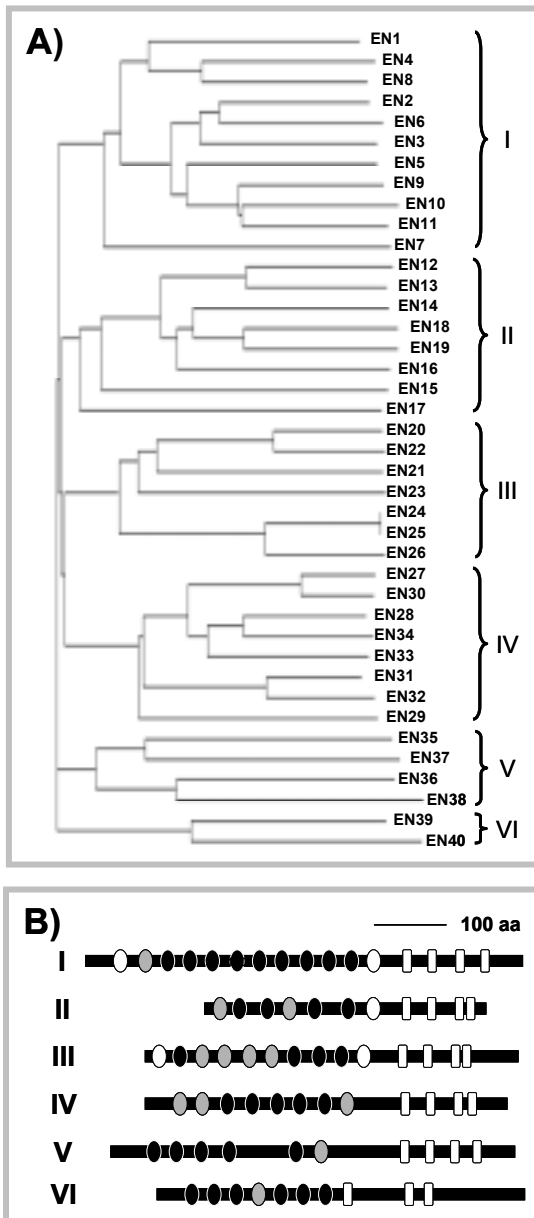


Figura 19.- Familias de proteínas de Arabidopsis que contienen repeticiones anquirina y dominios transmembrana (ANKTM). (A) Árbol filogenético de la región N-terminal de las proteínas (región de repeticiones anquirina) mediante método *Neighbor-joining*. Los números romanos a la derecha indican las diferentes familias (I a VI). (B) Representación esquemática de la proteína consenso para cada una de las familias ANKTM. Los círculos representan las repeticiones anquirina. Los círculos negros representan repeticiones anquirina presentes en >90% de las proteínas de la familia, los círculos grises entre 50 – 89% de las proteínas y los círculos blancos presentes en un porcentaje <50%. Los rectángulos vacíos representan los dominios transmembrana. Los números a la izquierda indican las diferentes familias (I a VI).

### II.2.1.2 Genes *AnkTm* en otras especies vegetales

La presencia de genes que codifican proteínas del tipo ANKTM en los genomas de otras especies vegetales fue estudiada mediante comparaciones de secuencia basadas en el programa TBLASTN (TIGR, <http://tigrblast.tigr.org/tgi/>). Cada una de las secuencias AtANKTM se comparó con la base de secuencias de plantas. Las secuencias obtenidas se tradujeron a proteína y se analizó la presencia de dominios conocidos mediante el servidor SMART v 3.5. Estos análisis han identificado genes que codifican proteínas ANKTM en especies de diferentes familias (Poaceae, Solanaceae ó Leguminosae), clases (Magnoliopsida (dicotiledóneas) y Liliopsida (monocotiledóneas)) e incluso divisiones (gimnospermas). Entre las especies podemos citar la cebolla (*Allium cepa*), el algodón (*Gossypium* sp.), la vid (*Vitis vinifera*), la lechuga (*Lactuca sativa*), el girasol (*Helianthus annuus*), el tomate (*Lycopersicon esculentum*), la petunia (*Petunia hybrida*), el álamo (*Populus trichocarpa*) y el abeto (*Picea glauca*). Entre todas las especies destaca el arroz (*Oryza sativa* L.), que es otro de los vegetales cuyo genoma ha sido secuenciado casi completamente (Yu *et al.*, 2002; Goff *et al.*, 2002), y que, sin embargo, solo cuenta con siete genes que codifican proteínas similares a las codificadas por los genes *AtAnkTm*. Cada una de estas nuevas secuencias se comparó con todas las identificadas en Arabidopsis y se determinó con cual de las seis familias de Arabidopsis guardaban mayor similitud (Cuadro 9). Sólo se han encontrado similitudes máximas para dos de las seis familias de Arabidopsis, la IV y la VI. Por ejemplo, de los siete genes de arroz, cuatro muestran mayores similitudes con la familia IV y tres con la familia VI.

Cuadro 9.- Genes *AnkTm* identificados en otras especies vegetales (30/10/05).

Código	Especie	Familia <i>AtAnkTm</i>	TC o GeneBank	Esquema
Os01	<i>Oryza sativa</i>	IV	AP003914	
Os02	<i>Oryza sativa</i>	IV	TC274663	
Os03	<i>Oryza sativa</i>	IV	TC252941	
Os04	<i>Oryza sativa</i>	IV	TC251163	
Os05	<i>Oryza sativa</i>	VI	TC264153	
Os06	<i>Oryza sativa</i>	VI	TC277294	
Os07	<i>Oryza sativa</i>	VI	TC265189	
Gm01	<i>Glicine max</i>	VI	TC219018	
Hv01	<i>Hordeum vulgare</i>	IV	TC140408	
Mt01	<i>Medicago truncatula</i>	VI	TC107689	
Pt01	<i>Pinus tadea</i>	IV	TC61335	
So01	<i>Saccharum officinarum</i>	IV	TC59217	
So02	<i>Saccharum officinarum</i>	IV	TC50269	
St01	<i>Solanum tuberosum</i>	VI	TC113700	
Ta01	<i>Triticum aestivum</i>	IV	TC237549	
Ta02	<i>Triticum aestivum</i>	IV	TC268176	
Zm01	<i>Zea mays</i>	IV	TC251100	

A continuación se estudió la presencia de genes *AnkTm* en *Brassica napus* (colza) una especie muy cercana a *Arabidopsis thaliana*. Esta especie de interés agronómico pertenece a la misma familia de *Arabidopsis thaliana* (Brassicaceae) y es una de las plantas de la que más ESTs se han depositado en las bases de secuencias (72.350; 10/01/06). Se diseñaron pares de oligonucleótidos iniciadores correspondientes a 13 genes de *Arabidopsis* y con ellos se realizaron amplificaciones por PCR a partir de DNA genómico tanto de *Arabidopsis* como de colza. En todos los casos se amplificaron bandas del tamaño esperado para *Arabidopsis* y se amplificaron bandas de similar tamaño en colza. Esto indica un alto grado de conservación de estas familias de genes entre ambas especies (Figura 20).

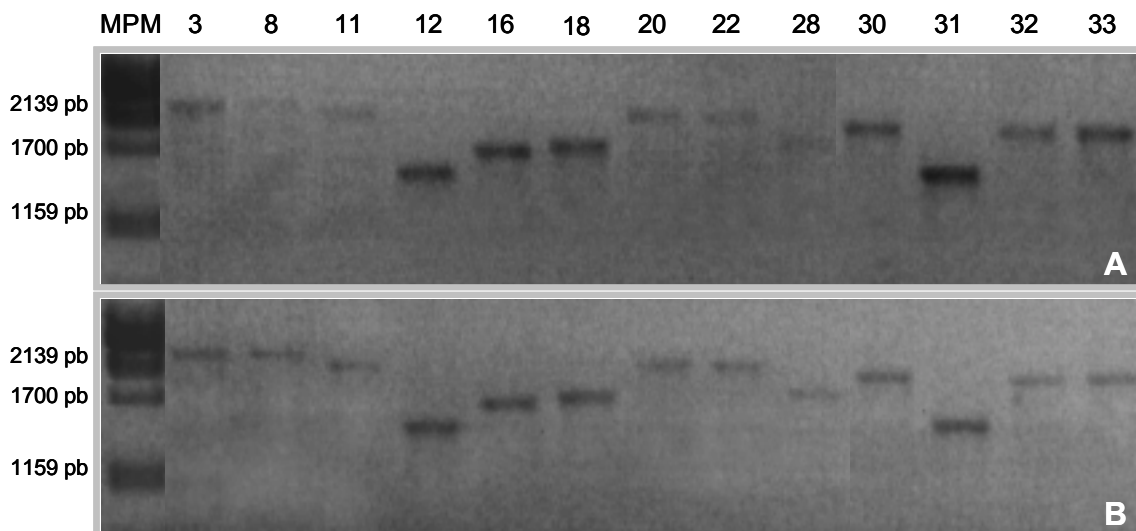


Figura 20.- Amplificación por PCR de fragmentos de 13 genes *AtAnkTm* (3, 8, 11, 12, 16, 18, 20, 22, 28, 30, 31, 32 y 33) en *Arabidopsis* y colza. (A) colza (*Brassica napus*); (B) *Arabidopsis thaliana*. Números a la izquierda de la figura indican el tamaño de los fragmentos de DNA del marcador de peso molecular (MPM) Lambda digerido con la enzima de restricción Pst<sub>I</sub>. Números en la parte superior indican el gen según el código *AtAnkTm*.

Por otra parte, se hizo una búsqueda de secuencias de EST correspondientes a genes *AnkTm* en colza utilizando el programa BLAST de TIGR (<http://tigrblast.tigr.org/tgi/>). Se encontraron dos secuencias de ESTs semejantes a los genes 4, 8, 14, 18, 28 y 29; una secuencia de EST semejante a los genes 1, 6, 11, 12, 17, 19, 21, 25, 27, 33 y 34. Las familias de genes *AtAnkTm* de mayor representación en cuanto número de secuencias

depositadas para *Brassica napus* son la IV y VI (12 y 4 ESTs, respectivamente).

## II.2.2. Organización genómica de los genes *AtAnkTm*

### II.2.2.1. Distribución de exones e intrones

Se analizó la distribución de intrones de todos los genes *AtAnkTm*. De ellos sólo un gen no tiene intrones (26). La gran mayoría de los intrones (89 de 96) están localizados en las regiones que codifican las repeticiones ANK, sin embargo, la posición de los exones no está correlacionada con la posición de las repeticiones ANK como sí se ha encontrado para algunos genes en mamíferos y vegetales (Albert *et al.*, 1999), así, 41 de los intrones interrumpen la región codificante para una repetición ANK. Se analizó tanto el número de intrones/exones como su distribución para comprobar si están relacionados con la filogenia. Algunos intrones están presentes en la misma posición en más de un gen de la misma familia, pero no en todos los casos (Figura 21).

### II.2.2.2. Distribución cromosómica

Los genes *AtAnkTm* están distribuidos en todos los cromosomas, aunque no uniformemente. Mientras los cromosomas II y III tienen cuatro genes cada uno, el cromosoma V tiene 13. Es posible encontrar algunas áreas con alta densidad de genes, tales como la parte inferior del cromosoma V y la superior del cromosoma I (Figura 22), así como también, amplias regiones que están desprovistas de genes *AtAnkTm*, éstas incluyen la parte superior de los cromosomas I y IV.

Hay cinco casos de dos o más genes agrupados en tándem (Figura 22). Tres de ellos corresponden a parejas de genes (10 y 11, 12 y 13, y 18 y 19), uno incluye tres genes (24, 25 y 26) y finalmente, hay 7 genes en tándem en el cromosoma IV (2 a 8) (Figura 23).

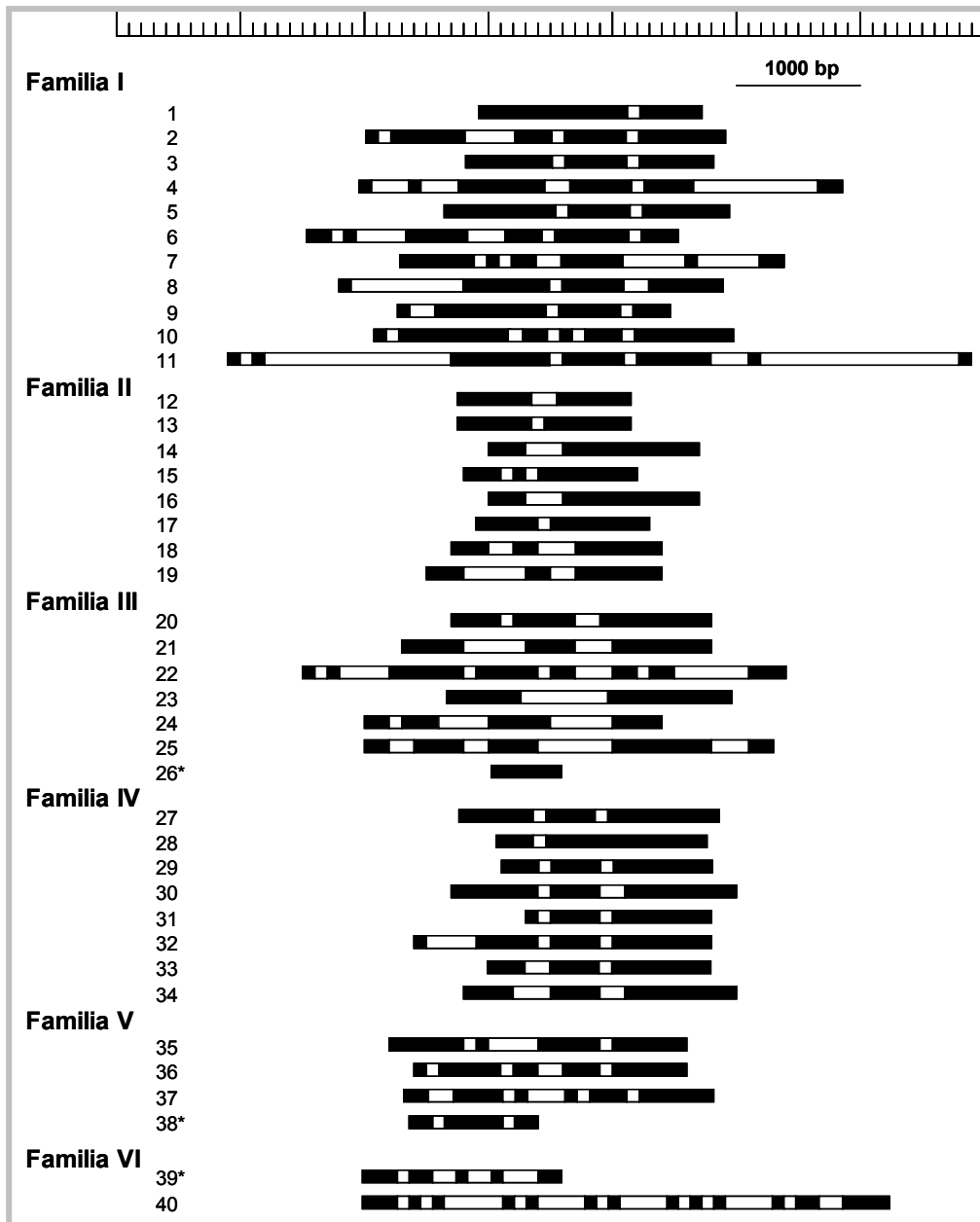
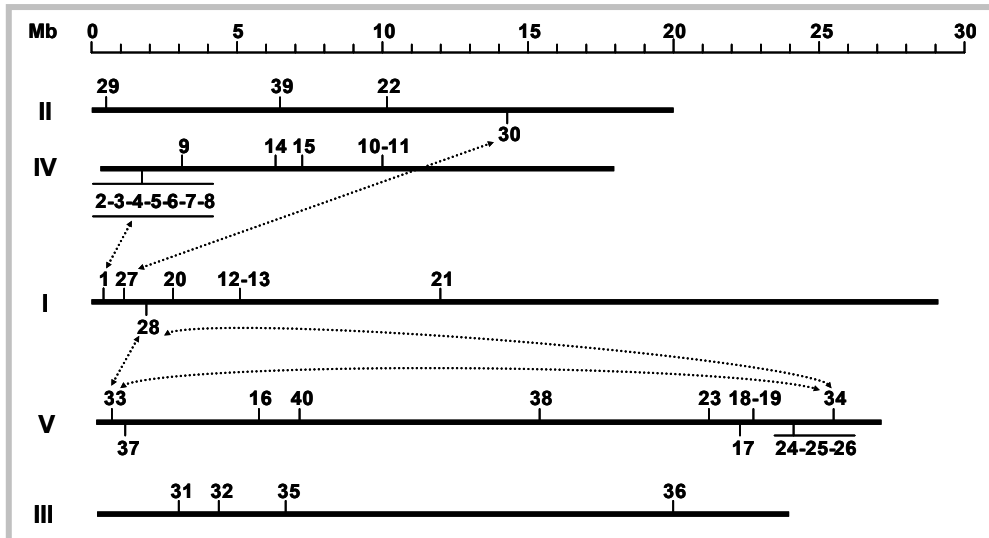


Figura 21.- Distribución de exones e intrones en los genes *AnkTm* de *Arabidopsis thaliana*. Rectángulos negros y blancos representan exones e intrones, respectivamente. \* indican los genes que codifican proteínas sin dominios transmembrana.



40

Figura 22.- Distribución cromosómica y eventos de duplicación de los genes *AtAnkTm* de *Arabidopsis thaliana*. Posiciones cromosómicas deducidas de los genes *AtAnkTm*. La escala es en Mb. Números separados por guiones representan genes repetidos en tándem. Líneas punteadas indican una correlación entre la región genómica duplicada y la presencia de genes *AtAnkTm*. Números I a V corresponden a cada uno de los cromosomas.

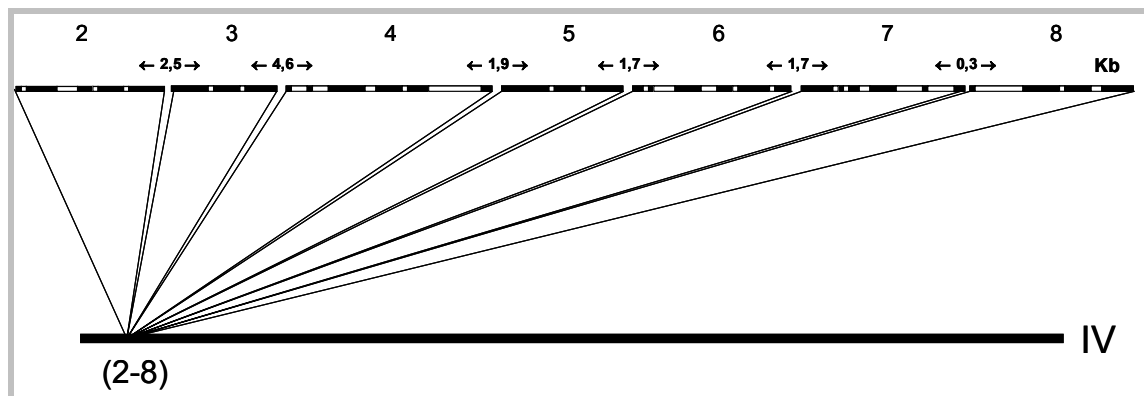


Figura 23.- Distribución de los genes *AtAnkTm* 2 a 8 en el cromosoma IV de *Arabidopsis thaliana*. Número entre flechas indica la separación que existe entre cada uno de los genes, expresado en Kb. Rectángulos negros y blancos señalan la distribución exon e intrón de cada uno de ellos, respectivamente.

El análisis de las áreas duplicadas inter- e intracromosómicas del genoma de *Arabidopsis* indica que hay tres casos de correlación entre la localización de genes y duplicaciones de genoma (Figura 22). El gen 27 en el cromosoma I parece estar duplicado en el cromosoma II (30). Seguramente por esta razón los genes 27 y 30 están muy relacionados en el análisis filogenético (Figura 19.A). Los genes 28, 33 y 34 están localizados en un área repetida tres veces en el genoma de *Arabidopsis*, dos en el cromosoma V y una en el cromosoma I. También estos genes están estrechamente relacionados de



acuerdo al análisis filogenético. Finalmente, el gen 1 localizado en una región del cromosoma I, está duplicado en el área del cromosoma IV que contiene el grupo de 7 genes *AtAnkTm* en tándem (Figura 22). En el análisis filogenético el gen 1 está muy relacionado con los genes 4 y 8 (Figura 19.A).

### II.2.3 Expresión de los genes *AtAnkTm*.

El análisis del patrón de expresión de los genes *AtAnkTm* se realizó mediante RT-PCR (Reverse Transcription – Polymerase Chain Reaction) semicuantitativa. Los resultados fueron comparados con datos de expresión obtenidos *in silico*: abundancia de ESTs en las bases de secuencias y análisis de bases de resultados de hibridaciones de micromatrices. El procedimiento seguido para cada una de las técnicas de análisis de la expresión utilizadas se describe a continuación.

- *RT-PCR semicuantitativa*. Se realizaron ensayos con dos tipos de muestras:
  - 1.- RNAs extraídos de órganos de plantas adultas crecidas en condiciones normales:
    - Raíz (Rz)
    - Hoja de roseta (Hr)
    - Hoja caulinar (Hc)
    - Flor (Fl)
    - Tallo (Ta)
    - Silicuas en estadíos diferentes de desarrollo. Estos ensayos se realizaron con dos grupos de muestras. El primero se dividió en tres estadíos:
      - Inmaduras ( $S_1$ ): 1-7 días después de antesis (DDA)
      - Intermedias ( $S_2$ ): 8-14 DDA
      - Maduras ( $S_3$ ): 15-21 DDA

El segundo grupo se dividió en cinco estadíos:

- $S_I$ , 0 – 4 DDA
- $S_{II}$ , 4 – 8 DDA
- $S_{III}$ , 8 – 12 DDA
- $S_{IV}$ , 12 – 16 DDA

- $S_V$ , 17 – 21 DDA

2.- RNAs extraídos de plántulas de 10 días crecidas *in vitro* y sometidas durante 24 horas a diferentes condiciones de estrés:

- Calor (42°C)
- Frío (4°C)
- Estrés osmótico (Manitol, 440 mM)
- Salinidad (NaCl, 150 mM)
- Herida (hojas pinzadas)
- Oscuridad
- Anoxia (sumergidas en agua)
- Sequía (sin medio de cultivo)
- Estrés oxidativo (peróxido de hidrógeno, 10mM)
- Radiación ultravioleta (15 minutos, 1.5kJ/m<sup>2</sup>h)

Para poder detectar las amplificaciones artefactuales de DNAg se diseñaron iniciadores de manera que flanqueaban uno, o varios intrones. Para obtener resultados semicuantitativos se limitó el número de ciclos para así mantener la amplificación dentro de la fase exponencial. Se emplearon iniciadores para el gen de la actina como control de carga.

- *Análisis de ESTs*: Se realizó una búsqueda de *ESTs* en las bases de secuencias, clasificándose según los órganos de los que se obtuvieron o de las condiciones de crecimiento (normal o estrés) a que fueron sometidas las plantas.

- *Análisis de hibridaciones de micromatrices*: Se analizó el patrón de expresión basándose en la base de datos de micromatrices *Gene Expression Visualization* (GEV) de *AtGenExpress* (<http://www.weigelworld.org/resources/microarray/AtGenExpress/>). Se

obtuvieron datos comparables de la intensidad de las hibridaciones obtenidas en diferentes ensayos de micromatrices en diferentes tejidos de la planta y en distintas fases de desarrollo (Schmid *et al.*, 2005). Los genes 1, 4, 5, 13, 14, 16, 21, 25, 27, 28 y 37 no fueron analizados porque no están incluidos en los chips de Arabidopsis de Affymetrix. Los órganos analizados en estos ensayos (eje X de las figuras) se indican en el Cuadro 10.

Cuadro 10.- Muestras para las que se han realizado análisis por micromatrices y que se incluyen en los datos obtenidos por GEV (*Gene Expression Visualization de AtGenExpress* <http://www.weigelworld.org/resources/microarray/AtGenExpress/>).

Muestra	Genotipo	Edad (días)
Apice		
Apice caulinar, vegetativo + hojas jóvenes	Wt	7
Apice caulinar, vegetativo	Wt	7
Apice caulinar, transición	Wt	14
Apice caulinar, inflorescencia	Wt	21
Apice caulinar, inflorescencia	clv3-7	>21
Apice caulinar, inflorescencia	lfy-12	>21
Apice caulinar, inflorescencia	ap1-15	>21
Apice caulinar, inflorescencia	ap2-6	>21
Apice caulinar, inflorescencia	ufo-1	>21
Apice caulinar, inflorescencia	ap3-6	>21
Apice caulinar, inflorescencia	ag-12	>21
Órganos florales		
Pedicelos estadio 15	Wt	>21
Sépalos estadio 12	Wt	>21
Sépalos estadio 15	Wt	>21
Pétalos estadio 12	Wt	>21
Pétalos estadio 15	Wt	>21
Estambres estadio 12	Wt	>21
Estambres estadio 15	Wt	>21
Polen maduro	Wt	42
Carpelo estadio 12	Wt	>21
Carpelo estadio 15	Wt	>21
Flores		
Estadio 9	Wt	>21
Estadio 10/11	Wt	>21
Estadio 12	Wt	>21
Estadio 12	clv3-7	>21
Estadio 12	lfy-12	>21
Estadio 12	ap1-15	>21
Estadio 12	ap2-6	>21
Estadio 12	ufo-1	>21
Estadio 12	ap3-6	>21
Estadio 12	ag-12	>21
Estadio 15	Wt	>21
Fior	Wt	28
Hoja		
Cotiledones	Wt	7
Hojas 1 + 2	Wt	7
Hoja de roseta 4, 1 cm de longitud	Wt	10
Hoja de roseta 4, 1 cm de longitud	gl1-T	10
Hoja de roseta 2	Wt	17
Hoja de roseta 4	Wt	17
Hoja de roseta 6	Wt	17
Hoja de roseta 8	Wt	17
Hoja de roseta 10	Wt	17
Hoja de roseta 12	Wt	17
Hoja de roseta 12	gl1-T	17
Peciolo hoja 7	Wt	17
Zona proximal hoja 7	Wt	17
Zona distal hoja 7	Wt	17
Hoja	Wt	15
Hojas senescentes	Wt	35
Hojas caulinares	Wt	>21

Muestra	Genotipo	Edad (días)
Raíz		
Raíz	Wt	7
Raíz	Wt	17
Raíz	Wt	15
Raíz	Wt	8
Raíz	Wt	8
Raíz	Wt	21
Raíz	Wt	21
Semillas	Genotipo	Edad (días)
Silicuas (embrión estadio globular a corazón)	Wt	56
Silicuas (embrión estadio corazón)	Wt	56
Silicuas (embrión estadio corazón a torpedo)	Wt	56
Semillas (embrión estadio torpedo)	Wt	56
Semillas (embrión estadio <i>walking-stick</i> )	Wt	56
Semillas (embrión estadio <i>early curled cotyledon</i> )	Wt	56
Semillas (embrión estadio <i>curled cotyledon</i> )	Wt	56
Semillas (embrión estadio <i>green cotyledon</i> )	Wt	56
Tallo		
Hipocotilo	Wt	7
Primer nudo	Wt	>21
Segundo internudo	Wt	>21
Planta completa		
Plántula, parte aérea	Wt	7
Plántula, parte aérea	Wt	8
Plántula, parte aérea	Wt	8
Plántula, parte aérea	Wt	21
Plántula, parte aérea	Wt	21
Roseta tras la floración	Wt	21
Roseta tras la floración	Wt	22
Roseta tras la floración	Wt	23
Roseta antes de floración	Wt	7
Roseta antes de floración	Wt	14
Roseta antes de floración	Wt	21

● *Análisis fenotípico de líneas de inserción.* En los casos en que fue posible, se obtuvieron plantas de Arabidopsis con alguna inserción en los genes *AtAnkTm* y se observó si dicha inserción producía algún cambio en el fenotipo en condiciones normales de cultivo. Las líneas de inserción se obtuvieron a partir de diversas colecciones de acceso público: colección SAIL (Syngenta), SM (*Exon trapping Insert Consortium*), Salk (The Salk Institute for Biological Studies) y WiscDsLox (University of Wisconsin). En la mayoría de casos no se observaron fenotipos asociados a la inserción. Únicamente dos de las líneas

mostraron un fenotipo de letalidad en homocigosis y se tratarán en más profundidad en el siguiente apartado.

### II.2.3.1 Genes control

En los experimentos de RT-PCR semicuantitativa se utilizaron como controles de carga ampliaciones realizadas con oligonucleótidos iniciadores correspondientes al gen *actina* cuya expresión es constitutiva. Como era esperable, los resultados obtenidos para este gen muestran ampliaciones similares en todas las muestras analizadas tanto en órganos en condiciones normales (Figura 24) como en condiciones de estrés (Figura 25). Los resultados del análisis mediante GEV muestran una intensidad de expresión del gen de *actina* similar en casi todos los órganos (Figura 26).

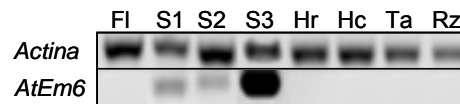


Figura 24.- Amplificaciones por RT-PCR de los genes control. FI, flores; S1, silicuas inmaduras; S2, silicuas intermedias; S3, silicuas maduras; Hr, hoja de roseta; Hc, hoja caulinar; Ta, tallo; Rz, raíz.

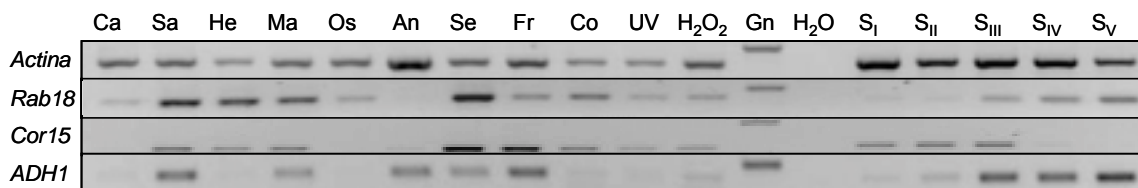


Figura 25.- Amplificaciones por RT-PCR de los genes control en plantas sometidas a diferentes tratamientos de estrés y en silicuas a diferentes estados de desarrollo. Tratamientos: calor (Ca), sal (Sa), herida (He), manitol (Ma), oscuridad (Os), anoxia (An), sequía (Se), frío (Fr), luz ultravioleta (UV) y peróxido de hidrogeno (H<sub>2</sub>O<sub>2</sub>). Controles: Co, plantas control no tratadas; Gn, DNA genómico; H<sub>2</sub>O, control de reacción de RT-PCR sin cDNA; S<sub>I</sub> representa a silicuas de 0 – 4 días después de polinización (DDP); S<sub>II</sub>, 4 – 8 DDP; S<sub>III</sub>, 8 – 12 DDP; S<sub>IV</sub>, 12 – 16 DDP; S<sub>V</sub>, 17 – 21 DDP.

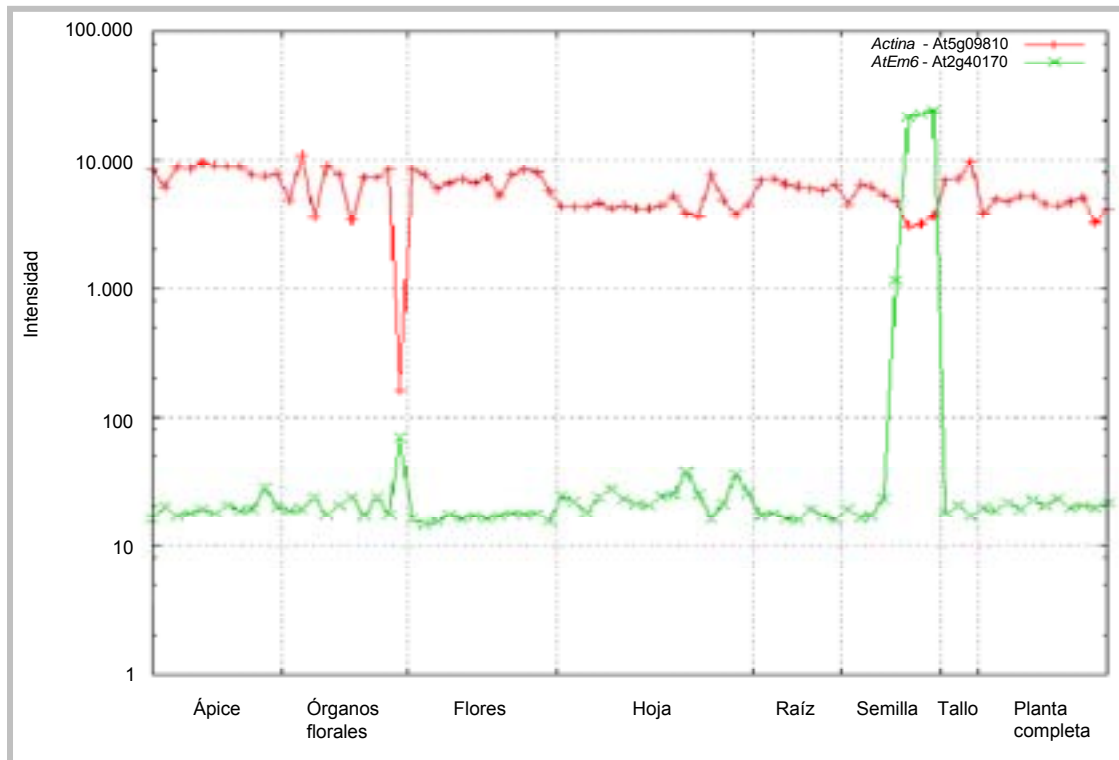


Figura 26.- Patrones de expresión de los genes *actina* y *AtEm6* basados en datos de hibridación de micromatrices. Eje X, diferentes órganos vegetales. Eje Y, valores de intensidad de hibridación. Cada punto de la gráfica corresponde a la media de tres intensidades de hibridación. Datos obtenidos a partir de AtGenExpress, *Gene Expression Visualization* (GEV) (<http://www.weigelworld.org/resources/microarray/AtGenExpress/>) (Schmid *et al.*, 2005).

Un segundo control utilizado fue el gen *AtEm6*, un gen que se transcribe específica y abundantemente durante la maduración de semilla (Vicent *et al.*, 2000). La amplificación por PCR usando oligonucleótidos para *AtEm6* estuvo acorde con lo esperado (Figura 24). Los resultados obtenidos en GEV también están de acuerdo con un gen cuya expresión es específica de etapas finales de desarrollo de semilla (Figura 26).

Se utilizaron otros controles para comprobar que los tratamientos de estrés se habían realizado correctamente. Se utilizaron oligonucleótidos diseñados para la amplificación específica de distintos genes para los que se ha descrito que se inducen en diferentes condiciones de estrés. Concretamente:

- *rab18* (Responsive to ABA 18; At5g66400). Se ha descrito que se induce en condiciones de sequía y frío, y en respuesta a ácido abscísico (Mantyla *et al.*, 1995)
- *Cor15A* (Cold-regulated 15A; At2g42540). Se ha descrito que se induce en tratamientos de frío y en respuesta a ABA (Wilhelm y Thomashow, 1993).
- *ADH1* (Alcohol Dehidrogenase 1; At1g77120). Se ha descrito que se induce por sequía, frío y anoxia (Dolferus *et al.*, 1994).

Los resultados de las amplificaciones por RT-PCR para estos genes son concordantes con lo previamente publicado. El gen *rab18* se induce por sequía, salinidad, herida y estrés osmótico, y esta reprimido por anoxia. El gen *Cor15* se induce principalmente por sequía y frío, y algo menos por salinidad, herida, manitol, luz ultravioleta y estrés oxidativo. El gen *ADH1* se transcribe bajo tratamiento salino y frío y también existe amplificación de bandas para los tratamientos de manitol, anoxia y sequía (Figura 25).

#### II.2.3.2 Genes de la familia I

Esta familia contiene 11 genes entre los cuales se encuentra el único gen *AtAnkTm* previamente caracterizado (11) y que codifica la proteína ACD6, que es un posible regulador y efector de la señal del ácido salicílico en la respuesta de defensa (Lu *et al.*, 2003). Las ESTs depositados en las bases de secuencias correspondientes a genes de esta familia se muestran en el cuadro 11. Los resultados de las amplificaciones por RT-PCR correspondientes a los genes de esta familia se muestran en las figuras 27 y 28. Los resultados de las hibridaciones de micromatrices obtenidos a partir de GEV se representan en la figura 29.

Cuadro 11.- ESTs correspondientes a la familia I de genes *AtAnkTm* (12-11-05)

Gen	Número Atg	Número de ESTs	Descripción de la genoteca
1	At1g03670	No	-
2	At4g03440	1	Planta completa
		1	Diferentes estados, deshidratación-frío
		1	Deshidratación-Rehidratación
3	At4g03450	1	Raíces (4 - 7 semanas)
		1	Planta completa
		1	Hojas, tratamiento de ozono
4	At4g03460	No	-
5	At4g03470	No	-
6	At4g03480	No	-
7	At4g03490	1	Planta completa
8	At4g03500	1	Flores y silicuas
		1	Roseta, deshidratación
9	At4g05040	2	Botones florales y flores
		1	Flores y silicuas
		3	Roseta, frío
		2	Oscuridad
		4	Varios tipos de estrés y hormonas
10	At4g14390	No	-
11	At4g14400 (ACD6)	1	Plántulas
		1	Parte aérea (2 - 6 semanas)
		1	Roseta (4 - 7 semanas)
		23	Roseta, frío
		2	Roseta, deshidratación
		1	Hojas, inducción por ozono
		1	Hojas infectadas con <i>Peronospora parasitica</i>
		3	Hojas infectadas con <i>Erysiphe cichoracearum</i> , plantas de 3 semanas
		5	
		3	Distintos estados, deshidratación y frío
7	Distintos tejidos y tratamientos		
			Varios tipos de estrés y hormonas

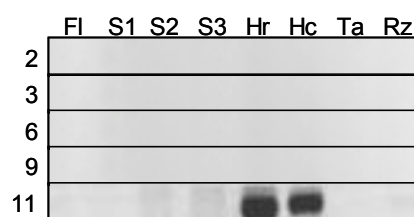


Figura 27.- Amplificaciones por RT-PCR de la familia I de genes *AtAnkTm*. El número que corresponde a cada gen está a la izquierda de cada fila. FI, flores; S1, silicuas inmaduras; S2, silicuas intermedias; S3, silicuas maduras; Hr, hoja de roseta; Hc, hoja caulinar; Ta, tallo; Rz, raíz.

- Resultados -

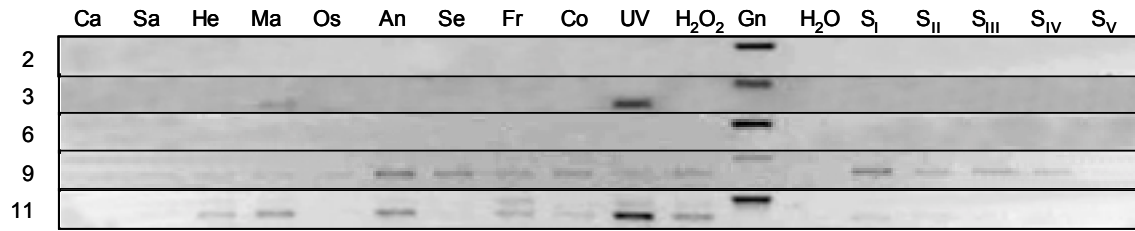


Figura 28.- Amplificaciones por RT-PCR de genes de la familia I de *AtAnkTm* en plantas sometidas a diferentes tratamientos de estrés y en silicuas a diferentes estados de desarrollo. El número que corresponde a cada gen está a la izquierda de cada fila. Tratamientos: calor (Ca), sal (Sa), herida (He), manitol (Ma), oscuridad (Os), anoxia (An), sequía (Se), frío (Fr), luz ultravioleta (UV) y peróxido de hidrogeno ( $H_2O_2$ ). Controles: Co, plantas control no tratadas; Gn, DNA genómico;  $H_2O$ , control de reacción de RT-PCR sin cDNA;  $S_I$  representa a silicuas de 0 – 4 días después de antesis (DDA);  $S_{II}$ , 4 – 8 DDA;  $S_{III}$ , 8 – 12 DDA;  $S_{IV}$ , 12 – 16 DDA;  $S_V$ , 17 – 21 DDA.

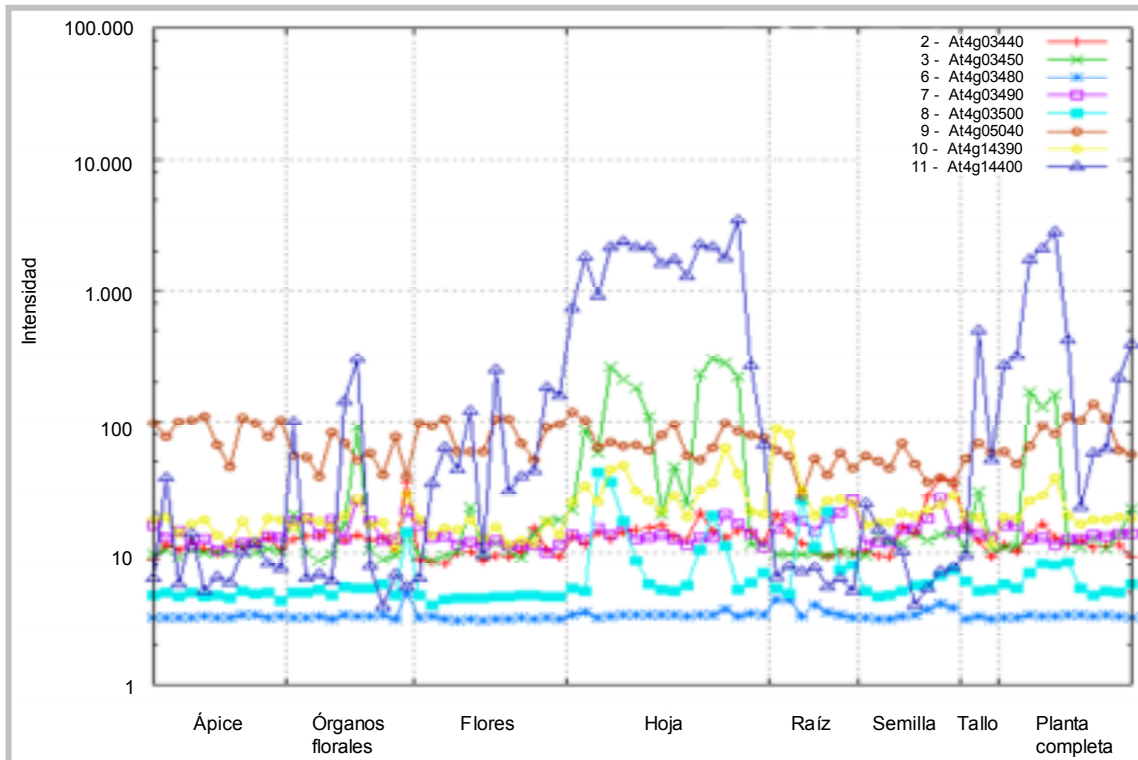


Figura 29.- Patrones de expresión de genes de la familia I de *AtAnkTm* basados en datos de hibridación de micromatrices. Eje X, diferentes órganos vegetales. Eje Y, valores de intensidad de hibridación. Cada punto de la gráfica corresponde a la media de tres intensidades de hibridación. Datos obtenidos a partir de AtGenExpress, *Gene Expression Visualization* (GEV) (<http://www.weigelworld.org/resources/microarray/AtGenExpress/>) (Schmid *et al.*, 2005).



**AtAnkTm1.** No se han encontrado ESTs ni se ha estudiado por RT-PCR. Tampoco está presente en las micromatrices de Affymetrix, por lo que no hay datos disponibles en GEV. Se analizaron las líneas Salk\_070803 y Salk\_095446, con una inserción en la región codificante del gen 1, y no se observó fenotipo en condiciones normales de desarrollo.

**AtAnkTm2.** Existen tres ESTs en las bases de secuencias correspondientes a este gen. Dos de ellos se obtuvieron de plantas sometidas a estrés por deshidratación-frío y por deshidratación. No se detectó amplificación de banda para el gen 2 en experimentos de RT-PCR en ninguno de los diferentes órganos estudiados bajo condiciones normales de crecimiento de la planta, ni en las diferentes condiciones de estrés. De igual forma, los datos obtenidos del GEV muestran que la intensidad de expresión es baja en todos los órganos estudiados, pero se incrementa algo en semilla a estadíos intermedios de desarrollo y en polen. La línea SAIL\_63 con una inserción en la región codificante muestra letalidad en homocigosis y se estudiará en más detalle en el siguiente capítulo.

**AtAnkTm3.** Se han secuenciado tres ESTs correspondientes a este gen. No se encontró amplificación de banda mediante RT-PCR en los diferentes órganos en condiciones normales de desarrollo, pero sí en los tratamientos de luz ultravioleta y, con menor intensidad, de estrés por manitol. Utilizando GEV se observaron valores relativamente altos de expresión en hojas y sépalos. La línea SM\_3\_31790, que posee una inserción en la región codificante, no presenta fenotipo diferente del silvestre.

**AtAnkTm4.** No hay ESTs. No se estudió por RT-PCR. No se encuentra en las micromatrices de Affymetrix. Únicamente se han estudiado las líneas SM\_3\_21515 y SM\_3\_2835, con una inserción en la región codificante, pero no se observó fenotipo alterado.

**AtAnkTm5.** No se poseen datos de expresión para este gen.

**AtAnkTm6.** No existen ESTs en las bases de secuencias y tampoco se ha encontrado amplificación por RT-PCR en condiciones normales ni en

ninguna de las condiciones de estrés estudiadas. Los valores de intensidad de expresión de GEV son muy bajos en todos los órganos. Las líneas Salk\_007058 y Salk\_063147, con inserciones en uno de los intrones y en la región codificante, respectivamente, no presentan fenotipo aparente bajo condiciones normales de desarrollo.

**AtAnkTm7.** Hay un EST correspondiente a este gen en las bases de secuencias, pero corresponde a una genoteca obtenida por mezcla de diversos tejidos, por lo que no es demasiado informativo. Los valores de intensidad de hibridación son bajos para este gen al estudiar el patrón de expresión por GEV.

**AtAnkTm8.** Se han encontrado dos ESTs correspondientes a este gen. Los valores de intensidad de expresión son bajos para el gen 8 al estudiar el patrón de expresión por GEV. Para el gen 8 se estudió la línea mutante WiscDsLox442C8, con una inserción en la región codificante, no observándose diferencias de fenotipo.

**AtAnkTm9.** Se han encontrado doce ESTs correspondientes a este gen en las bases de secuencias correspondientes a genotecas de silicua y flor, o en respuesta a varios estreses. Los experimentos de RT-PCR muestran que este gen se transcribe en silicuas en estadíos tempranos de desarrollo, y en plántulas, observándose inducción en respuesta a anoxia y, algo menos, a sequía. Su expresión se reprime en respuesta a varios tratamientos (calor, salinidad, herida, manitol, ultravioleta y oscuridad). Los valores obtenidos en GEV son similares para todos los órganos estudiados. La línea SAIL\_642 con una inserción en la región codificante muestra letalidad en homocigosis y se estudiará en más detalle en el siguiente capítulo.

**AtAnkTm10.** No presenta ESTs en las bases de secuencias y no se estudió por RT-PCR. El gen 10 no posee valores de intensidad de hibridación muy altos por medio de GEV, pero se observa cierta mayor hibridación en raíces y hojas de plantas de pocos días.

**AtAnkTm11.** Este gen es claramente el que posee la mayor intensidad de expresión de todos los estudiados. Por un lado, se han recopilado hasta 48 secuencias de ESTs correspondientes a este gen, la mayor parte obtenidas a partir de hojas sometidas a distintos tipos de estrés o infección. Los experimentos de RT-PCR muestran una amplificación muy intensa en hojas. En silicuas, se encontró expresión muy débil en los estadios iniciales del desarrollo. Este resultado es consistente con el patrón de expresión señalado por Lu y colaboradores (2003). Los resultados de RT-PCR muestran también un incremento importante de la intensidad de la amplificación en plantas sometidas a luz ultravioleta, anoxia, manitol, herida, frío y peróxido de hidrógeno. Los resultados de RT-PCR son similares a los observados en GEV en el que existe una alta intensidad de hibridación especialmente en hoja, tallo y sépalos.

### II.2.3.3 Genes de la familia II

Esta familia contiene ocho genes y se han secuenciado ESTs para cinco de ellos (Cuadro 12). Se ha estudiado la expresión de todos ellos por RT-PCR (Figuras 30 y 31), pero sólo están incluidas sondas para cinco de ellos en las micromatrices de Affymetrix (Figura 32).

Cuadro 12.- ESTs correspondientes a la familia I de genes *AtAnkTm* (12-11-05)

Gen	Número Atg	Número de ESTs	Descripción de genoteca
12	At1g14480	2 3	Raíces Hipocotilo, cultivo de tejidos; cicloheximida
13	At1g14500	No	
14	At4g10720	1	Raíces
15	At4g11000	No	
16	At5g15500	5 1 1	Flores y silicuas Hojas de plántulas de 2 a 3 semanas; Tratamiento de sales Deshidratación – Rehidratación
17	At5g51160	1 2	Raíces 4 – 7 semanas Varios tipos de estrés y hormonas
18	At5g54610	2 1	Deshidratación – Rehidratación Varios tipos de estrés y hormonas
19	At5g54620	No	

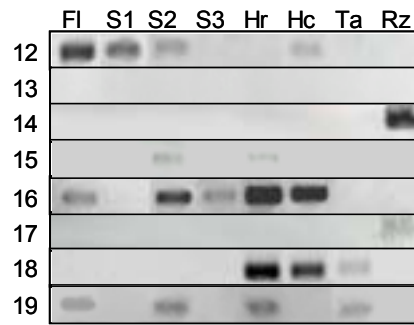


Figura 30.- Amplificaciones por RT-PCR de la familia II de genes *AtAnkTm*. El número que corresponde a cada gen está a la izquierda de cada fila. FI, flores; S1, silicuas inmaduras; S2, silicuas intermedias; S3, silicuas maduras; Hr, hoja de roseta; Hc, hoja caulinar; Ta, tallo; Rz, raíz.

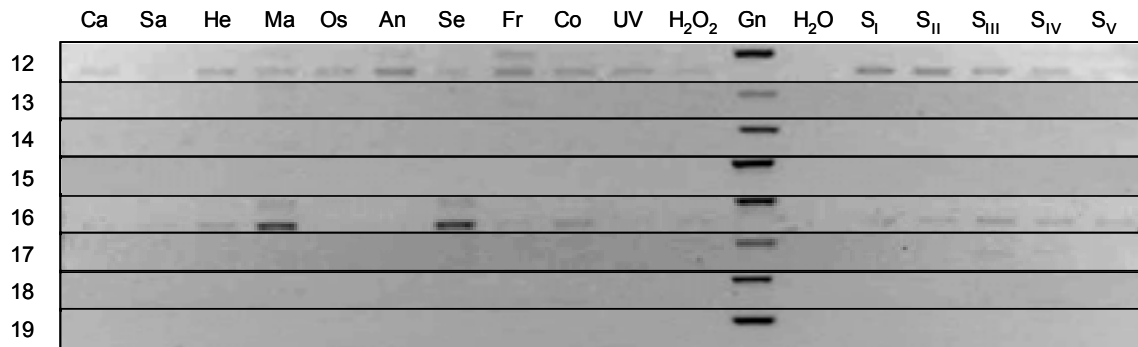


Figura 31.- Amplificaciones por RT-PCR de genes de la familia II de *AtAnkTm* en plantas sometidas a diferentes tratamientos de estrés y en silicuas a diferentes estados de desarrollo. El número que corresponde a cada gen está a la izquierda de cada fila. Tratamientos: calor (Ca), sal (Sa), herida (He), manitol (Ma), oscuridad (Os), anoxia (An), sequía (Se), frío (Fr), luz ultravioleta (UV) y peróxido de hidrogeno (H<sub>2</sub>O<sub>2</sub>). Controles: Co, plantas control no tratadas; Gn, DNA genómico; H<sub>2</sub>O, control de reacción de RT-PCR sin cDNA; S<sub>I</sub> representa a silicuas de 0 – 4 días después de anthesis (DDA); S<sub>II</sub>, 4 – 8 DDA; S<sub>III</sub>, 8 – 12 DDA; S<sub>IV</sub>, 12 – 16 DDA; S<sub>V</sub>, 17 – 21 DDA.

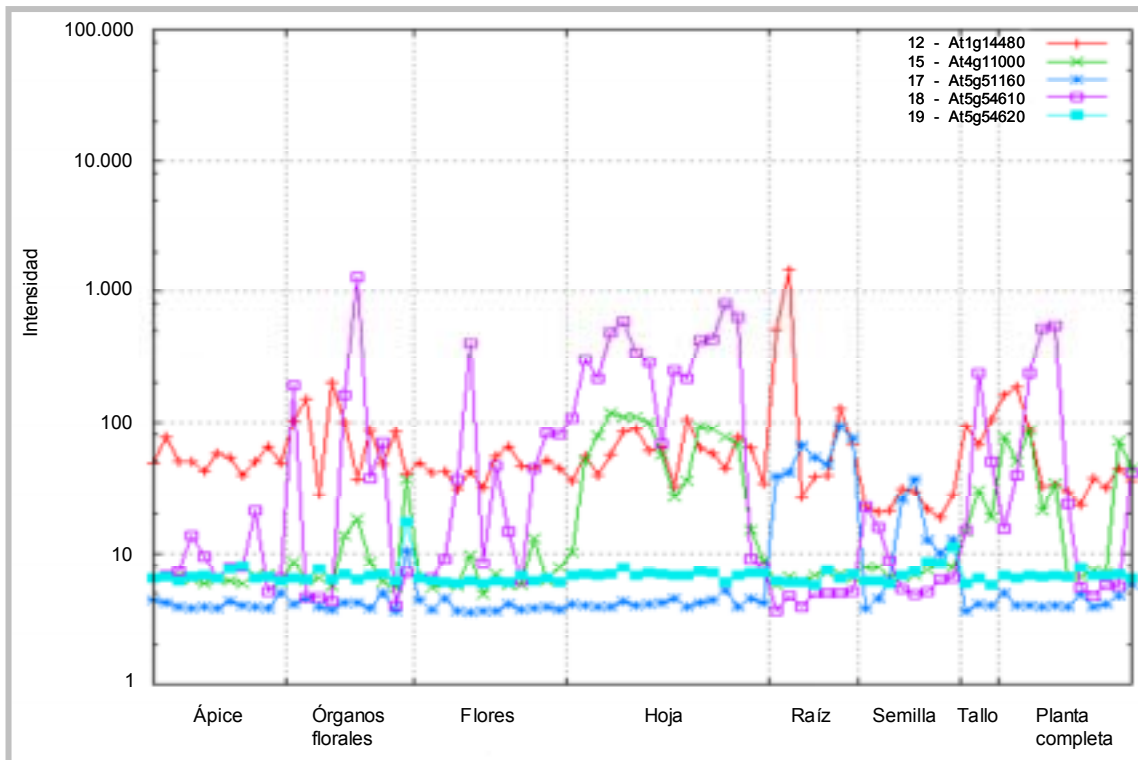


Figura 32.- Patrones de expresión de genes de la familia II de *AtAnkTm* basados en datos de hibridación de micromatrices. Eje X, diferentes órganos vegetales. Eje Y, valores de intensidad de hibridación. Cada punto de la gráfica corresponde a la media de tres intensidades de hibridación. Datos obtenidos a partir de AtGenExpress, *Gene Expression Visualization* (GEV) (<http://www.weigelworld.org/resources/microarray/AtGenExpress/>) (Schmid *et al.*, 2005).

***AtAnkTm12***. El gen 12 parece transcribe a niveles importantes en flores y silicuas en etapas tempranas de desarrollo. También se observa amplificación por RT-PCR en hojas caulinares, pero no de roseta. Al estudiar la transcripción en condiciones de estrés se encontró expresión en casi todos los tratamientos empleados. Mediante GEV, la intensidad de expresión es similar en casi todos los órganos de la planta excepto en raíz y en ciertos órganos florales. Se han encontrado cinco ESTs correspondientes a este gen.

***AtAnkTm13***. No se encontró amplificación por RT-PCR en ninguno de los órganos ni en las condiciones de estrés, excepto una leve señal en la muestra de plantas tratadas con frío. No existe información en la base de datos de hibridación de micromatrices respecto a la expresión de este gen ni se han secuenciado ESTs.

***AtAnkTm14***. Se amplifican bandas correspondientes a este gen específicamente en raíz, coincidiendo con lo encontrado en las bases de datos

de ESTs. No se encontró transcripción en ninguno de los tratamientos de estrés empleados. No existe información para este gen en GEV.

**AtAnkTm15.** La expresión es muy débil en silicua intermedia y en hoja y no se detecta transcripción en los tratamientos de estrés estudiados. Por GEV, los valores de intensidad más altos se encuentran en hojas y algo menores en tallo y algunos órganos florales. La línea SAIL\_44, que posee una inserción en la región codificante, no presenta fenotipo diferente del silvestre.

**AtAnkTm16.** El patrón de expresión del gen 16 indica que esta asociado a órganos reproductivos y hojas. Este gen parece inducirse por estrés debido a manitol y por sequía (Figura 31). No existe información de expresión para este gen en GEV. Se han secuenciado siete ESTs, cinco de las cuales corresponden a flores y silicuas. Las líneas Salk\_019783, Salk\_071042 y SAIL\_66, que poseen una inserción en la región codificante, la primera, y en la región del promotor las dos últimas, no presentaron fenotipo diferente del silvestre asociado a la inserción.

**AtAnkTm17.** Para el gen 17 se encontró expresión en raíz por RT-PCR y muy débil en silicua en estadíos intermedios de desarrollo. No fue posible detectar amplificación en las RT-PCR de diferentes condiciones de estrés. Mediante GEV se observaron mayores intensidades de hibridación de expresión en raíz y en estadíos intermedios de desarrollo de la semilla. La línea Salk\_084573, que posee una inserción en la región del promotor, no presenta fenotipo diferente del silvestre.

**AtAnkTm18.** Las amplificaciones de RT-PCR muestran resultados positivos en hojas y algo menos en tallo. No se observa amplificación en ninguna condición de estrés. Existen tres secuencias de EST correspondientes a este gen en las bases de secuencias. Los resultados de GEV señalan una elevada intensidad de hibridación en hojas, tallos y sépalos.

**AtAnkTm19.** Se detecta amplificación por RT-PCR en diversos órganos como flor, silicua intermedia, hoja y tallo, pero no en respuesta a ninguna de las condiciones de estrés estudiadas. Los valores de intensidad de hibridación de

GEV son bajos excepto para polen, que son algo mayores. Se analizó el fenotipo de la línea Salk\_053630, con una inserción en la región codificante, no encontrándose diferencias respecto al silvestre.

### II.2.3.4 Genes de la familia III

De los siete genes integrantes de esta familia, tres no tienen ESTs descritas hasta la fecha (Cuadro 13), se han realizado experimentos de RT-PCR en dos de ellos (Figuras 33 y 34) y cinco están incluidos en la micromatriz de Affymetrix (Figura 35).

Cuadro 13.- ESTs correspondientes a la familia III de genes *AtAnkTm* (12-11-05)

Gen	Número Atg	Número de ESTs	Descripción de genoteca
20	At1g10340	1	Flores y silicuas
		2	Tejido vegetativo adulto
		2	Hojas (Nº 5 y 6 a partir de la base) en período de senescencia
		3	Distintos estados; Deshidratación y frío
		2	Varios tipos de estrés y hormonas
		1	Hojas infectadas con <i>Peronospora parasitica</i>
		1	Hojas infectadas con <i>Erysiphe cichoracearum</i> , plantas de 3 semanas
21	At1g34050	No	
22	At2g24600	1	Raíces 4 – 7 semanas
		4	Roseta; frío
		1	Roseta; deshidratación
		2	Varios tipos de estrés y hormonas
		2	Tratamiento hormonal en callos
23	At5g50140	No	
24	At5g54700	No	
25	At5g54710	1	Parte aérea de 2 – 6 semanas
		2	Roseta; frío
		2	Hojas; inducción por ozono
26	At5g54720	1	Parte aérea de 2 – 6 semanas
		2	Hojas; inducción por ozono

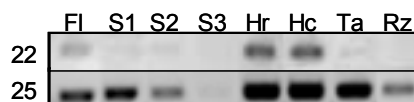


Figura 33.- Amplificaciones por RT-PCR de la familia III de genes *AtAnkTm*. El número que corresponde a cada gen está a la izquierda de cada fila. Fl, flores; S1, silicuas inmaduras; S2, silicuas intermedias; S3, silicuas maduras; Hr, hoja de roseta; Hc, hoja caulinar; Ta, tallo; Rz, raíz.

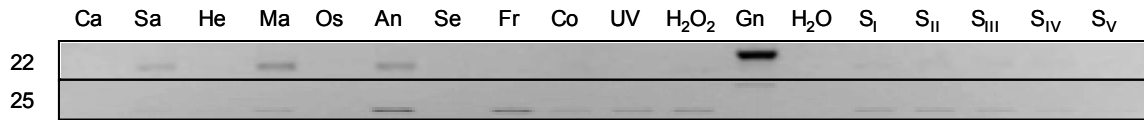


Figura 34.- Amplificaciones por RT-PCR de genes de la familia III de *AtAnkTm* en plantas sometidas a diferentes tratamientos de estrés y en silicuas a diferentes estados de desarrollo. El número que corresponde a cada gen está a la izquierda de cada fila. Tratamientos: calor (Ca), sal (Sa), herida (He), manitol (Ma), oscuridad (Os), anoxia (An), sequía (Se), frío (Fr), luz ultravioleta (UV) y peróxido de hidrogeno (H<sub>2</sub>O<sub>2</sub>). Controles: Co, plantas control no tratadas; Gn, DNA genómico; H<sub>2</sub>O, control de reacción de RT-PCR sin cDNA; S<sub>I</sub> representa a silicuas de 0 – 4 días después de antesis (DDA); S<sub>II</sub>, 4 – 8 DDA; S<sub>III</sub>, 8 – 12 DDA; S<sub>IV</sub>, 12 – 16 DDA; S<sub>V</sub>, 17 – 21 DDA.

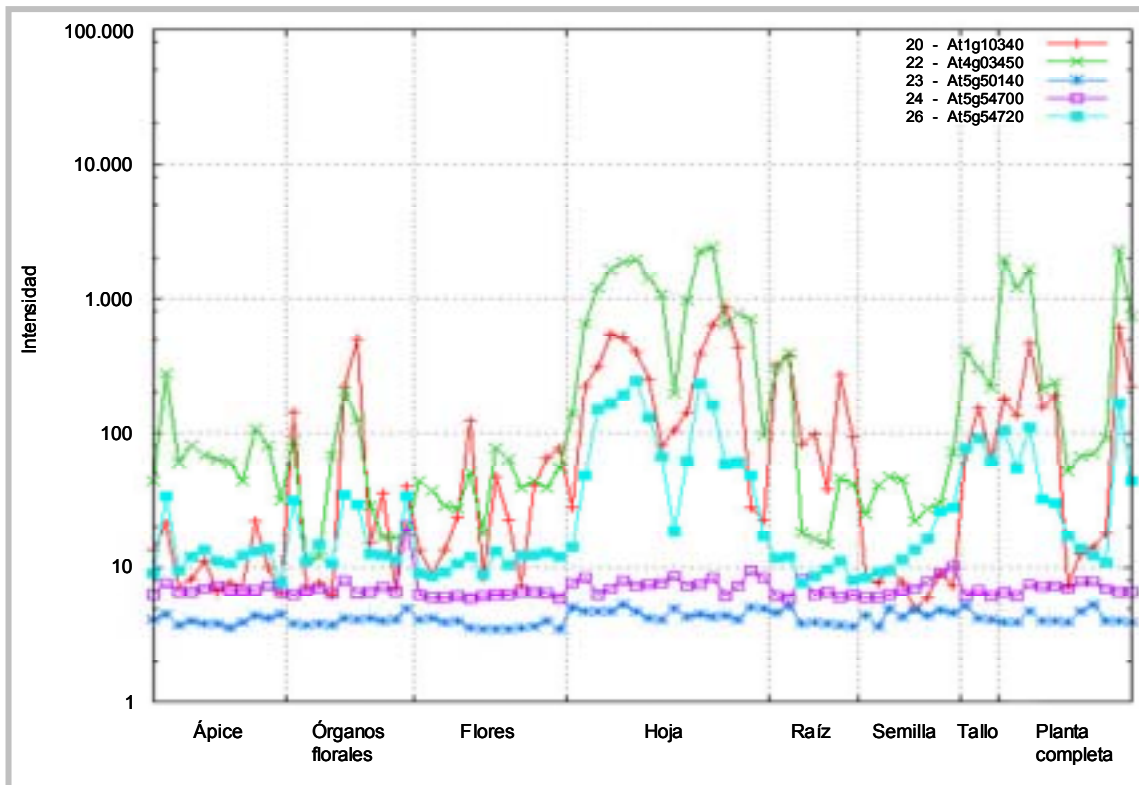


Figura 35.- Patrones de expresión de genes de la familia III de *AtAnkTm* basados en datos de hibridación de micromatrices. Eje X, diferentes órganos vegetales. Eje Y, valores de intensidad de hibridación. Cada punto de la gráfica corresponde a la media de tres intensidades de hibridación. Datos obtenidos a partir de AtGenExpress, *Gene Expression Visualization* (GEV) (<http://www.weigelworld.org/resources/microarray/AtGenExpress/>) (Schmid *et al.*, 2005).

***AtAnkTm20***. Los datos de hibridación de micromatrices referentes al gen 20 en GEV señalan que su intensidad de expresión es, en general, alta en todos los órganos a excepción de semillas y órganos florales que no sean sépalos. Se han encontrado doce ESTs correspondientes a este gen correspondientes, sobre todo, a genotecas a distintos tipos de estrés. Se analizó la línea Salk\_059400 con una inserción en la región codificante del gen



20 no encontrándose un fenotipo diferente del silvestre bajo condiciones normales de desarrollo.

**AtAnkTm21.** No existen datos de expresión para este gen.

**AtAnkTm22.** Los experimentos de RT-PCR indican que este gen se expresa en hojas y flores, y en respuesta a algunos tratamientos de estrés como salinidad, manitol y anoxia. Se han encontrado diez secuencias de ESTs correspondientes a este gen provenientes de raíces y diferentes tratamientos de estrés. Los resultados de las hibridaciones de micromatrices depositados en GEV indican que este gen se expresa especialmente en hoja, tallo y sépalos.

**AtAnkTm23.** No se han encontrado ESTs correspondientes a este gen. Las intensidades de hibridación de micromatrices para el gen 23 en GEV son muy bajas en todos los casos. Se analizaron las líneas SM\_3\_32888 y SM\_3\_32883, con una inserción en la región codificante del gen 23 y no se observó fenotipo alterado en condiciones normales de desarrollo.

**AtAnkTm24.** No se han encontrado ESTs correspondientes a este gen. Las intensidades de hibridación de micromatrices para el gen 24 en GEV son muy bajas en todos los casos excepto en polen.

**AtAnkTm25.** Los ensayos de RT-PCR muestran expresión en casi todos los órganos de la planta examinados excepto silicua madura, y en respuesta a algunos de los tratamientos de estrés, principalmente anoxia y frío. No existe información de hibridaciones de micromatrices en GEV para este gen. Se han encontrado cinco ESTs correspondientes al gen 25.

**AtAnkTm26.** La expresión del gen 26 por GEV indica que la más alta intensidad de expresión se encuentra en hojas, tallo, sépalos y polen. Existen tres secuencias de EST en las bases para este gen.

#### II.2.3.5 Genes de la familia IV

Esta familia contiene ocho genes. Para tres de estos genes (27, 28 y 34) no se ha identificado ESTs a la fecha (Cuadro 14). Se han realizado

experimentos de RT-PCR en seis de ellos (Figuras 36 y 37) y seis están incluidos en la micromatriz de Affymetrix (Figura 38).

Cuadro 14.- ESTs correspondientes a la familia IV de genes *AtAnkTm* (12-11-05)

Gen	Número Atg	Número de ESTs	Descripción de genoteca
27	At1g05640	No	
28	At1g07710	No	
29	At2g01680	1	Botones florales y flores
		2	Flores y silicuas
		3	Silicuas verdes
		2	Parte aérea, 2 a 6 semanas
		1	Raíces de 4 a 7 semanas
		1	Hojas (Nº 5 y 6 a partir de la base) en período de senescencia
		1	Mezcla de tejidos
		1	Planta de 6 semanas; distintos tratamientos
		3	Distintos estados; deshidratación y frío
		3	Varios tipos de estrés y hormonas
30	At2g31820	1	Raíces de 4 a 7 semanas
		1	Hojas de roseta de 4 semanas infectadas con <i>Pseudomonas syringae</i> pv. <i>tomato</i>
31	At3g09550	1	Roseta; frío
		2	Distintos tejidos y tratamientos
32	At3g12360	1	Plántulas
		1	Distintos tejidos y tratamientos
		1	Hojas infectadas con <i>Erysiphe cichoracearum</i> , plantas de 3 semanas
33	At5g02620	1	Botones florales
		4	Flores y silicuas
		2	Roseta; frío
		2	Crecimiento en oscuridad
		1	Frío
		2	Distintos estados; deshidratación y frío
		1	Varios tipos de estrés y hormonas
34	At5g60070	No	

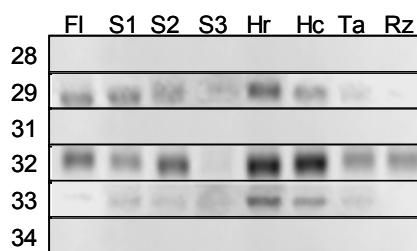


Figura 36.- Amplificaciones por RT-PCR de la familia IV de genes *AtAnkTm*. El número que corresponde a cada gen está a la izquierda de cada fila. FI, flores; S1, silicuas inmaduras; S2, silicuas intermedias; S3, silicuas maduras; Hr, hoja de roseta; Hc, hoja caulinar; Ta, tallo; Rz, raíz.

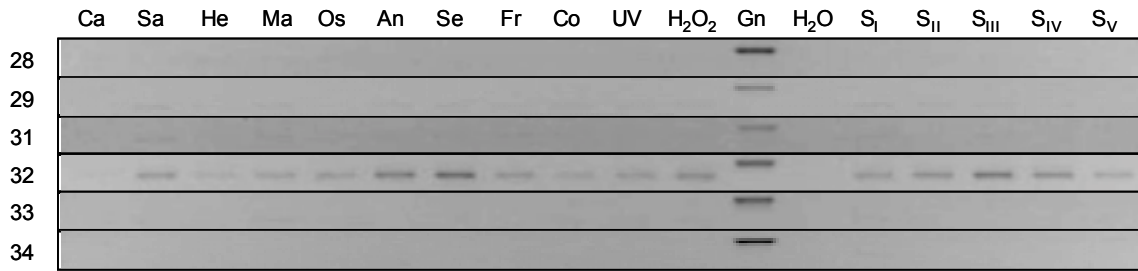


Figura 37.- Amplificaciones por RT-PCR de genes de la familia IV de *AtAnkTm* en plantas sometidas a diferentes tratamientos de estrés y en silicuas a diferentes estados de desarrollo. El número que corresponde a cada gen está a la izquierda de cada fila. Tratamientos: calor (Ca), sal (Sa), herida (He), manitol (Ma), oscuridad (Os), anoxia (An), sequía (Se), frío (Fr), luz ultravioleta (UV) y peróxido de hidrogeno (H<sub>2</sub>O<sub>2</sub>). Controles: Co, plantas control no tratadas; Gn, DNA genómico; H<sub>2</sub>O, control de reacción de RT-PCR sin cDNA; S<sub>I</sub> representa a silicuas de 0 – 4 días después de antesis (DDA); S<sub>II</sub>, 4 – 8 DDA; S<sub>III</sub>, 8 – 12 DDA; S<sub>IV</sub>, 12 – 16 DDA; S<sub>V</sub>, 17 – 21 DDA.

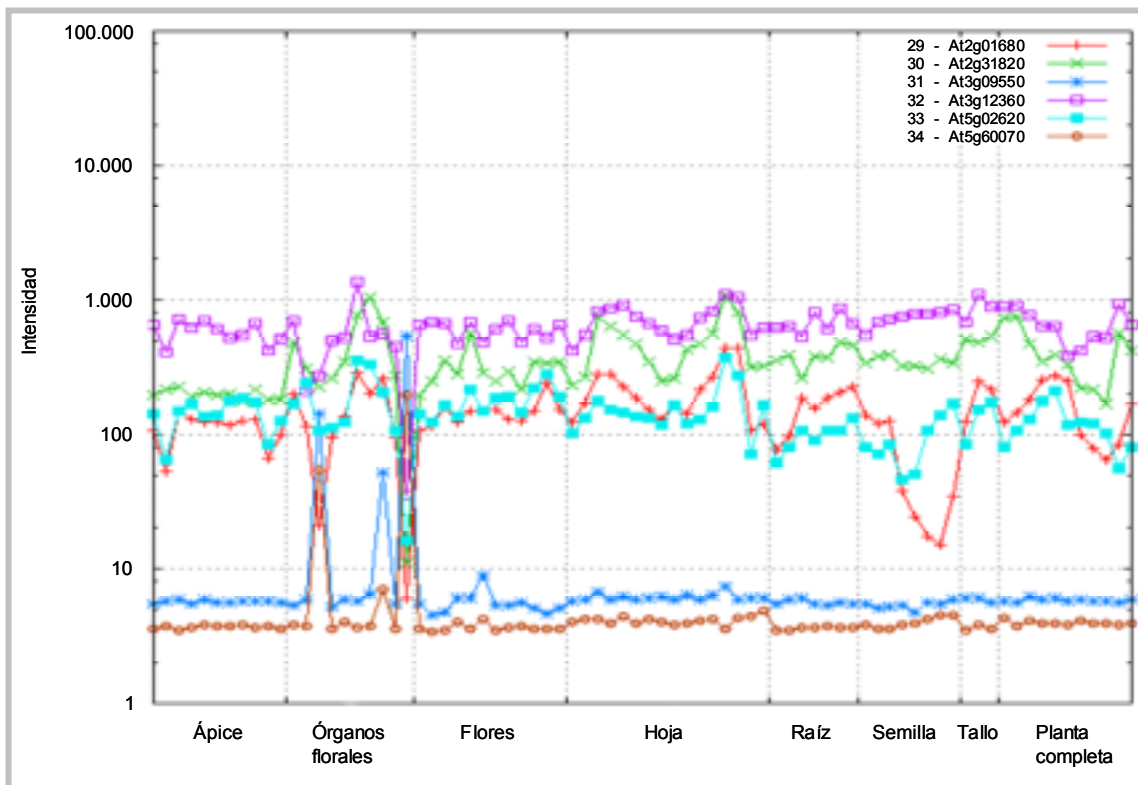


Figura 38.- Patrones de expresión de genes de la familia IV de *AtAnkTm* basados en datos de hibridación de micromatrices. Eje X, diferentes órganos vegetales. Eje Y, valores de intensidad de hibridación. Cada punto de la gráfica corresponde a la media de tres intensidades de hibridación. Datos obtenidos a partir de AtGenExpress, *Gene Expression Visualization (GEV)* (<http://www.weigelworld.org/resources/microarray/AtGenExpress/>) (Schmid *et al.*, 2005).

***AtAnkTm27***. No se poseen datos de expresión para este gen ni se ha localizado ningún EST. Se analizó la línea Salk\_007101 con una inserción en la región codificante del gen 27, no encontrándose fenotipo distinto del silvestre.

**AtAnkTm28.** No existe información respecto a este gen en GEV. No existen ESTs correspondientes a este gen. No se detectó expresión notable del gen 28 por RT-PCR en los órganos de la planta ni en las condiciones de estrés estudiadas. Únicamente se detectó una amplificación muy débil correspondiente a las etapas tempranas de desarrollo de la silicua que apenas es visible en las figuras 36 y 37. Esto se confirmó posteriormente, pero estos resultados serán comentados en un próximo capítulo. Las líneas Salk\_008522, Salk\_008523 y Salk\_043469, con una inserción en la región del promotor de este gen, no presentaron diferencias fenotípicas respecto del control.

**AtAnkTm29.** Se han identificado 21 ESTs correspondientes a este gen provenientes de distintos órganos de la planta y algunas condiciones de estrés. Los datos de RT-PCR indican que el gen 29 se expresa en casi todos los órganos estudiados, excepto raíz, y muy débilmente en diversas condiciones de estrés. Esto coincide con lo observado para los diferentes órganos por GEV en que la intensidad de hibridación es homogénea en los órganos analizados.

**AtAnkTm30.** El análisis de las micromatrices en GEV indica que este gen se expresa de manera uniforme y alta en los distintos órganos de la planta excepto en polen. Existen dos ESTs correspondientes a este gen en las bases de secuencias.

**AtAnkTm31.** Por RT-PCR no se encontró expresión del gen 31 en los diferentes órganos y sólo se observó una muy débil transcripción en los tratamientos de sal y manitol. Los datos de intensidad de hibridación en GEV muestran que el gen se transcribe a bajo nivel excepto en polen. Se han encontrado dos ESTs correspondientes a este gen. Se estudió la línea SAIL\_633 con una inserción en la región codificante de este gen y no fue posible identificar fenotipo diferente del control bajo condiciones normales de crecimiento.

**AtAnkTm32.** Mediante RT-PCR se encontró expresión en la mayoría de los órganos de la planta excepto semilla madura, y en algunos tratamientos de estrés como sequía, anoxia, peróxido de hidrógeno, frío y salinidad. Asimismo, mediante GEV la intensidad de expresión es alta en todos los órganos

estudiados excepto polen. Se han encontrado tres ESTs correspondientes a este gen.

**AtAnkTm33.** La amplificación de bandas de RT-PCR del gen 33 se puede apreciar principalmente en hojas y en menor grado en los otros órganos analizados excepto raíz, pero no en respuesta a estrés. Se han encontrado 13 ESTs correspondientes a diversos órganos y condiciones de crecimiento. La intensidad de hibridación en GEV es semejante y alta en todos los órganos excepto polen. Se analizó el fenotipo de la línea SAIL\_20, con una inserción en la región del promotor del gen 33, no observándose diferencias respecto al control bajo condiciones normales de crecimiento.

**AtAnkTm34.** Los resultados de las RT-PCR indican que el gen 34 no se expresa en los órganos estudiados ni en las condiciones de estrés. No se han encontrado ESTs correspondientes a este gen y además posee valores bajos de intensidad de hibridación en GEV excepto para el polen. Se analizaron las líneas SAIL\_140, SM\_3\_15526 y SM\_3\_15534, todas con inserción en la región codificante de este gen, no diferenciándose del fenotipo control.

#### II.2.3.6 Genes de la familia V

Esta familia contiene cuatro genes. Hasta la fecha, sólo se ha encontrado un EST en esta familia correspondiente al gen 37 (Cuadro 15). Solo se analizó por RT-PCR el gen 37 (Figuras 39 y 40). Existen datos en GEV para tres de los genes (Figura 41)

Cuadro 15.- ESTs correspondientes a la familia V de genes *AtAnkTm* (12-11-05)

Gen	Número Atg	Número de ESTs	Descripción de genoteca
35	At3g18670	No	
36	At3g54070	No	
37	At5g04690	1	Silicuas verdes
38	At5g35830	No	

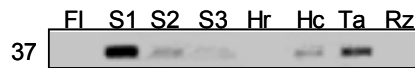


Figura 39.- Amplificaciones por RT-PCR de la familia V de genes *AtAnkTm*. El número que corresponde a cada gen está a la izquierda de cada fila. Fl, flores; S1, silicuas inmaduras; S2, silicuas intermedias; S3, silicuas maduras; Hr, hoja de roseta; Hc, hoja caulinar; Ta, tallo; Rz, raíz.

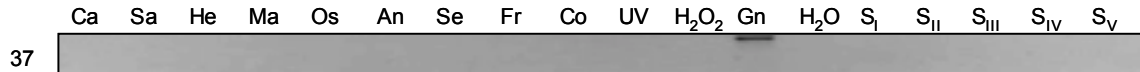


Figura 40.- Amplificaciones por RT-PCR de genes de la familia V de *AtAnkTm* en plantas sometidas a diferentes tratamientos de estrés y en silicuas a diferentes estados de desarrollo. El número que corresponde a cada gen está a la izquierda de cada fila. Tratamientos: calor (Ca), sal (Sa), herida (He), manitol (Ma), oscuridad (Os), anoxia (An), sequía (Se), frío (Fr), luz ultravioleta (UV) y peróxido de hidrogeno ( $H_2O_2$ ). Controles: Co, plantas control no tratadas; Gn, DNA genómico;  $H_2O$ , control de reacción de RT-PCR sin cDNA; S<sub>I</sub> representa a silicuas de 0 – 4 días después de antesis (DDA); S<sub>II</sub>, 4 – 8 DDA; S<sub>III</sub>, 8 – 12 DDA; S<sub>IV</sub>, 12 – 16 DDA; S<sub>V</sub>, 17 – 21 DDA.

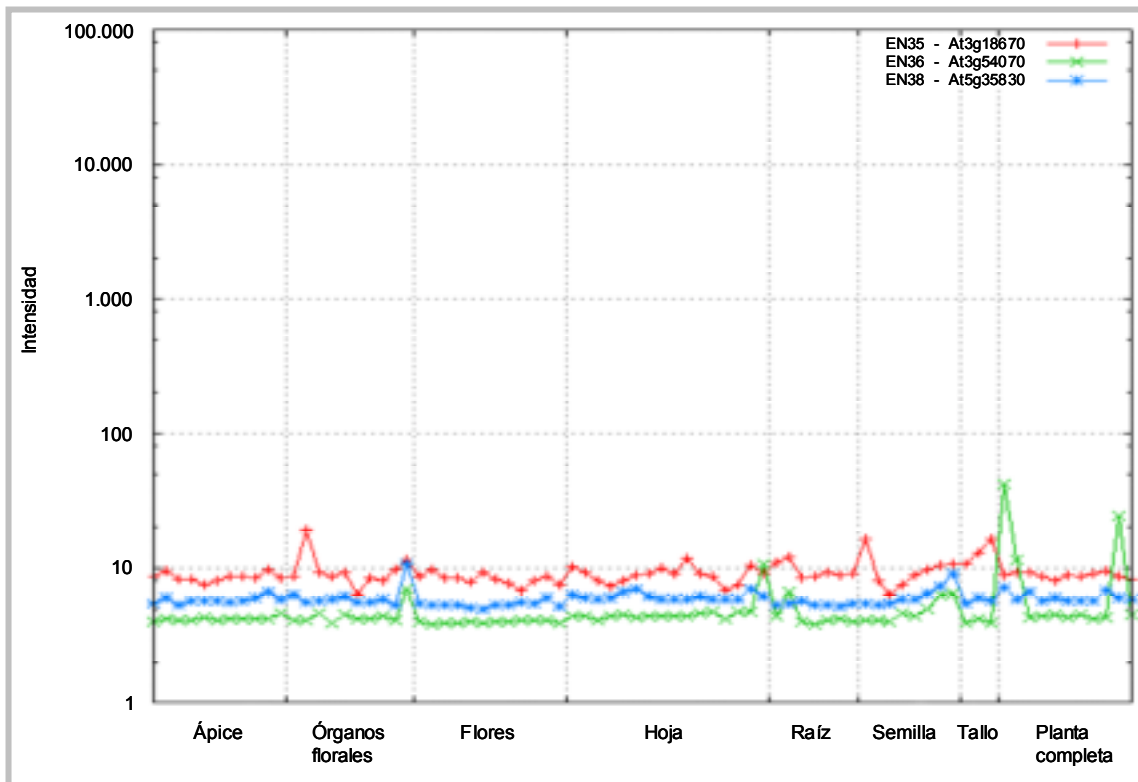


Figura 41.- Patrones de expresión de genes de la familia IV de *AtAnkTm* basados en datos de hibridación de micromatrices. Eje X, diferentes órganos vegetales. Eje Y, valores de intensidad de hibridación. Cada punto de la gráfica corresponde a la media de tres intensidades de hibridación. Datos obtenidos a partir de AtGenExpress, *Gene Expression Visualization* (GEV) (<http://www.weigelworld.org/resources/microarray/AtGenExpress/>) (Schmid *et al.*, 2005).

**AtAnkTm35.** No se conocen ESTs para este gen. Los datos en GEV indican que el gen tiene muy bajos niveles de expresión.

**AtAnkTm36.** No se conocen ESTs para este gen. Los datos en GEV indican que el gen tiene muy bajos niveles de expresión.

**AtAnkTm37.** Se ha encontrado un EST correspondiente a este gen obtenido a partir de silicuas inmaduras. Los datos de RT-PCR indican que el se expresa en tallo, hoja caulinar y en silicuas, especialmente en los estadíos iniciales de desarrollo. No se ha observado expresión en respuesta a estrés. No existe información para este gen en GEV. Se estudió la línea SM\_3\_16667, con inserción en la región codificante, pero no fue posible identificar diferencias en fenotipo con respecto al control.

**AtAnkTm38.** No existen ESTs secuenciados para este gen y los datos de GEV indican niveles muy bajos de expresión.

#### II.2.3.7 Genes de la familia VI

Esta familia contiene dos genes. No existen datos de RT-PCR para esta familia, pero si se han encontrado ESTs (Cuadro 16) y existen datos en GEV para ambos genes (Figura 42).

Cuadro 16.- ESTs correspondientes a la familia VI de genes *AtAnkTm* (12-11-05)

Gen	Número Atg	Número de ESTs	Descripción de genoteca
39	At2g14250	1	Distintos tejidos y tratamientos
40	At5g20350	2	Semillas de 5-13 días después de anthesis
		4	Flores y silicuas
		1	Silicuas verdes
		3	Raíces
		1	Meristemo floral, una semana después de emergido
		1	Parte aérea, 2 a 6 semanas
		1	Hojas (Nº 5 y 6 a partir de la base) en período de senescencia
		1	Mezcla de tejidos
		2	Distintos estados; deshidratación y frío
		2	Distintos tejidos y tratamientos
		1	Deshidratación – rehidratación
		3	Crecimiento en oscuridad
		2	Varios tipos de estrés y hormonas

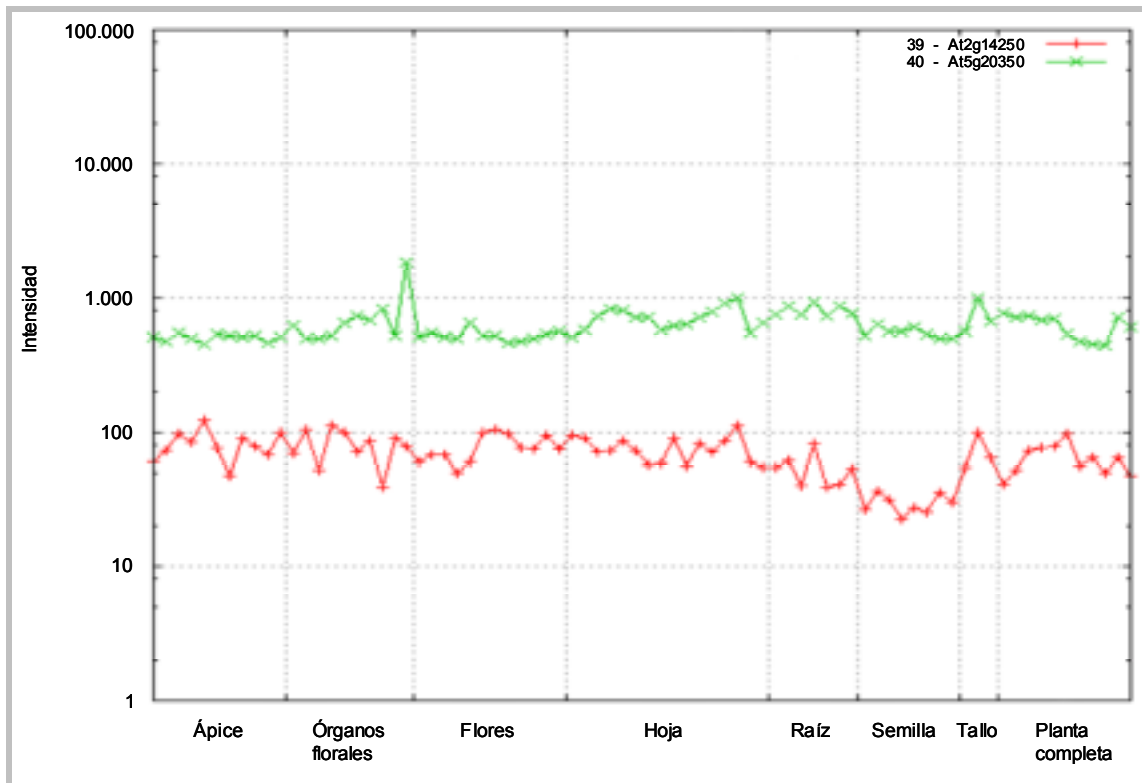


Figura 42.- Patrones de expresión de genes de la familia IV de *AtAnkTm* basados en datos de hibridación de micromatrices. Eje X, diferentes órganos vegetales. Eje Y, valores de intensidad de hibridación. Cada punto de la gráfica corresponde a la media de tres intensidades de hibridación. Datos obtenidos a partir de AtGenExpress, *Gene Expression Visualization* (GEV) (<http://www.weigelworld.org/resources/microarray/AtGenExpress/>) (Schmid *et al.*, 2005).

***AtAnkTm39***. Se ha encontrado un EST correspondiente a este gen. Los datos en GEV indican que este gen se expresa a un nivel uniforme en todos los órganos analizados.

***AtAnkTm40***. La intensidad de expresión del gen 40 es similar y alta entre los diferentes órganos de la planta, según GEV, siendo algo mayor en polen). Este alto nivel de hibridación coincide con el hecho de que se han encontrado 24 ESTs en los bancos de secuencias correspondientes a este gen.

#### II.2.4 Mutantes letales de la familia I

Como se ha comentado en el apartado anterior, únicamente dos de las líneas mutantes analizadas presentaban un fenotipo mutante en condiciones normales de cultivo. Las dos líneas fueron obtenidas de la colección SAIL de Syngenta (líneas SAIL 63 y 642) y corresponden a los genes 2 y 9 (Figura 43).



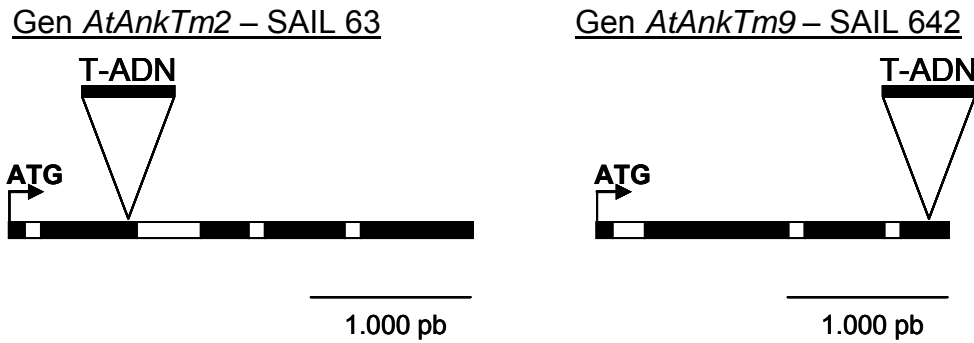


Figura 43.- Representación esquemática de las inserciones de T-DNA en los genes *AtAnkTm2* y *9*. Rectángulos negros corresponden a exones y rectángulos blancos a intrones.

Las plantas de las líneas mutantes se seleccionaron para la presencia de T-DNA con tratamientos con el herbicida Basta®. Se extrajo DNA genómico de 12 plantas supervivientes por línea. De cada planta, se realizaron dos amplificaciones por PCR, una para detectar la presencia del alelo salvaje y otra para detectar la inserción (Figura 44). Todas las plantas de las dos líneas resultaron ser heterocigotas. Por lo tanto, las inserciones en los genes *AtAnkTm2*, y *AtAnkTm9* producen letalidad en homocigosis.

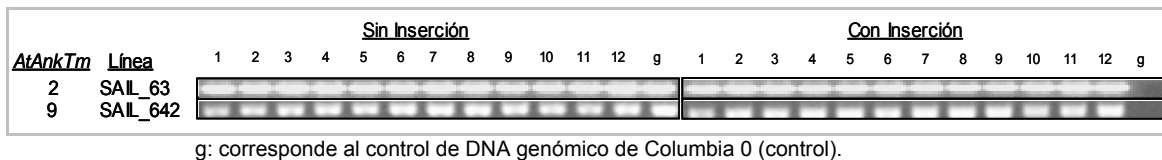


Figura 44.- Análisis a nivel de genotipo de líneas mutantes de genes de la Familia I *AtAnkTm* de *Arabidopsis thaliana*.

No se observaron alteraciones en el desarrollo de las semillas en ninguna de las dos líneas. Los porcentajes de germinación de las semillas fueron similares a los del ecotipo silvestre. Esto descartó que las alteraciones fuesen debidas a defectos en la embriogénesis.

Seguidamente, se estudió si las mutaciones afectaban a algún aspecto de la gametogénesis. Para ello, se realizaron una serie de experimentos histoquímicos mediante dos de los productos más utilizados en tinción para analizar los componentes celulares de los granos de polen. Uno de ellos es 4',6-diamina-2-fenilindola (DAPI), un colorante fluorescente que tiñe DNA de los

núcleos. Para la otra tinción se utilizó azul de anilina que contiene un fluorocromo (sirofluor) que se une específicamente a  $\beta$ -1,3-glucano, el principal componente de la pared del tubo del polen.

La tinción DAPI mostró que una proporción importante de los granos de polen de las dos líneas mutantes no teñían sus núcleos, a diferencia del control silvestre Columbia 0 en la que se tiñe casi la totalidad de los núcleos de los granos de polen (Figura 45). Los resultados obtenidos en la tinción DAPI para la línea SAIL\_642 (gen 9) son similares a los obtenidos en la línea SAIL\_63 (gen 2), por lo que sólo se muestran para dicha línea. Se contabilizaron los granos de polen teñidos de las flores mutantes de las líneas SAIL 63 y 642 y resultaron ser del 34 y 41% respectivamente. Estos valores fueron muy inferiores a los obtenidos para el control Columbia 0 (87 %).

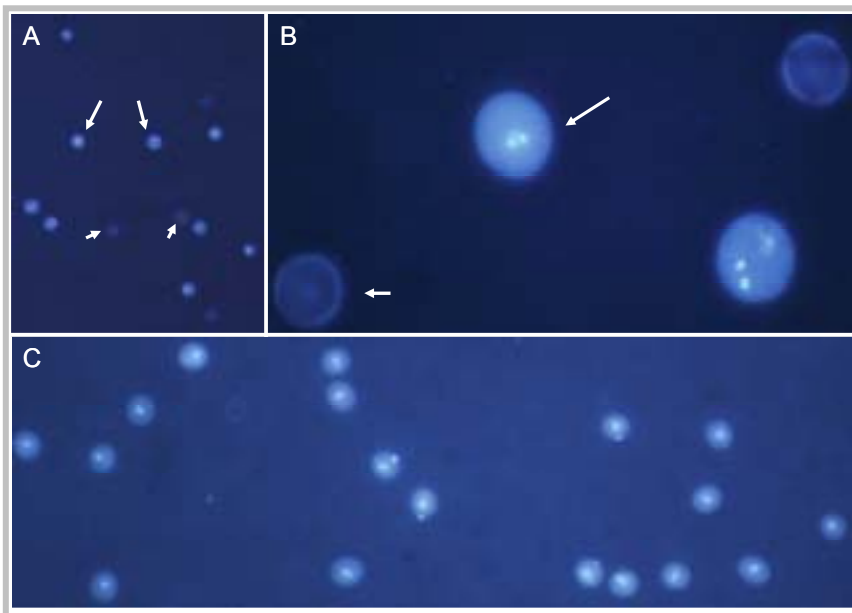


Figura 45.- Tinción DAPI de granos de polen de la línea SAIL\_642 (A y B) con una inserción de T-DNA en el gen *AtAnkTm9*. Se aprecian granos con los núcleos teñidos (flechas largas) y otros sin tinción (flechas cortas). En C, se puede apreciar la tinción de la casi totalidad de los núcleos de los granos de polen del control Columbia 0.

La tinción de azul de anilina también mostró diferencias entre las líneas mutantes y la silvestre como se puede apreciar en la figura 46. En el control se encontró una gran mayoría de los granos de polen con una intensidad de fluorescencia alta, respecto a las líneas mutantes que poseen una menor proporción de granos teñidos.

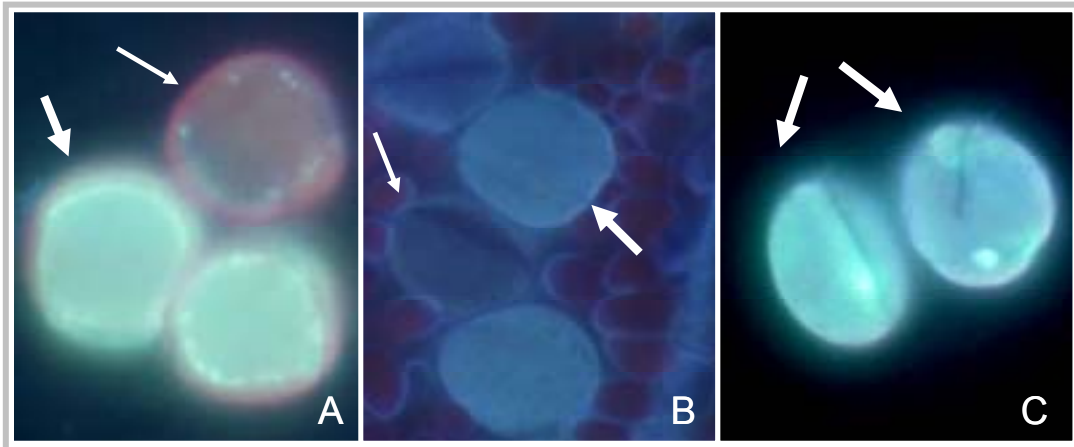


Figura 46.- Tinción con Azul de Anilina de los granos de polen de las líneas SAIL 63 (A) (Gen2) y 642 (B) (Gen9), y el control Columbia 0 (C). Granos de polen de tinción normal (flecha ancha) y con fluorescencia reducida (flecha angosta).

A continuación se empleó la microscopía electrónica de escaneo para comparar las características de la cubierta (exina) de los granos de polen de las líneas mutantes con respecto al control. Para ello, se separaron las anteras de las flores y se depositaron en filtros y los granos de polen fueron extraídos desde las anteras utilizando etanol absoluto. Mediante una reacción acetolítica se eliminó el contenido de los granos de polen dejando sólo la exina (Erdtman, 1960). Los granos de polen de la línea SAIL\_642 (Gen 9) (Figura 47.C y D) no presentan diferencias respecto al control (Figura 47.E y F). Si se observaron diferencias en el polen de la línea SAIL\_63 (Gen 2) (Figura 47.A y B). Los granos de polen de esta línea presentan un entramado más denso en la exina, dejando aperturas germinativas (espacios sin cubierta) de menor tamaño que el control.

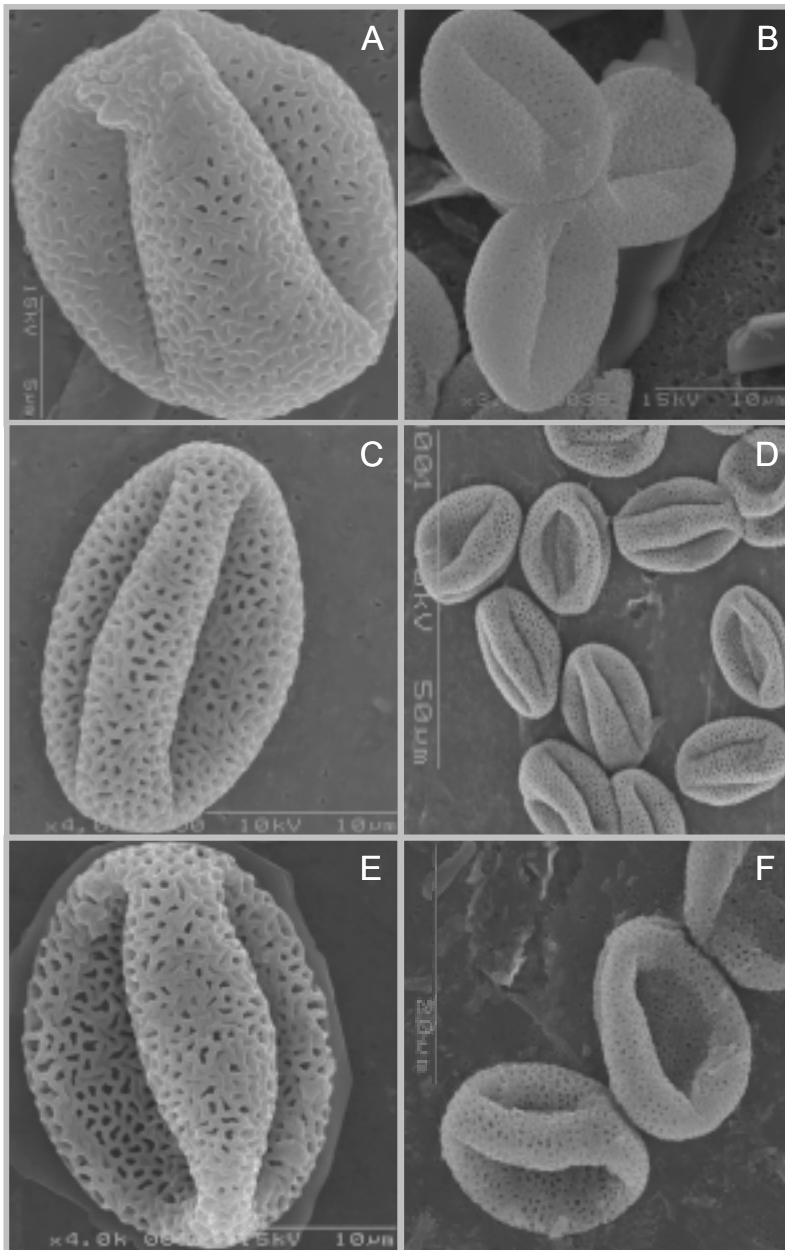


Figura 47.- Análisis de la cubierta del polen de las líneas mutantes y el control, Columbia – 0 mediante microscopio electrónico de barrido. A y B corresponden a la línea SAIL\_63 (Gen *AtAnkTm2*); C y D a la línea SAIL\_642 (Gen *AtAnkTm9*); E y F a la línea silvestre Columbia – 0. Barras, expresadas en  $\mu\text{m}$ .

### II.2.5 Análisis del gen *AtAnkTm28* de la familia IV.

El análisis de la expresión de los genes *AtAnkTm* mostró que el gen *AtAnkTm28* parecía tener un patrón de expresión específico de silicuas en etapas tempranas de desarrollo. Por ello se decidió profundizar en el estudio de este gen.

#### II.5.1 Representación esquemática del gen *AtAnkTm28*

El gen 28 posee dos exones de 343 y 1289 pb y un intrón de 80 pb (Figura 48). El RNA mensajero codifica una proteína de 543 aminoácidos, con

una región con ocho repeticiones ANK en N-terminal y una región con cuatro dominios transmembrana en C-terminal.

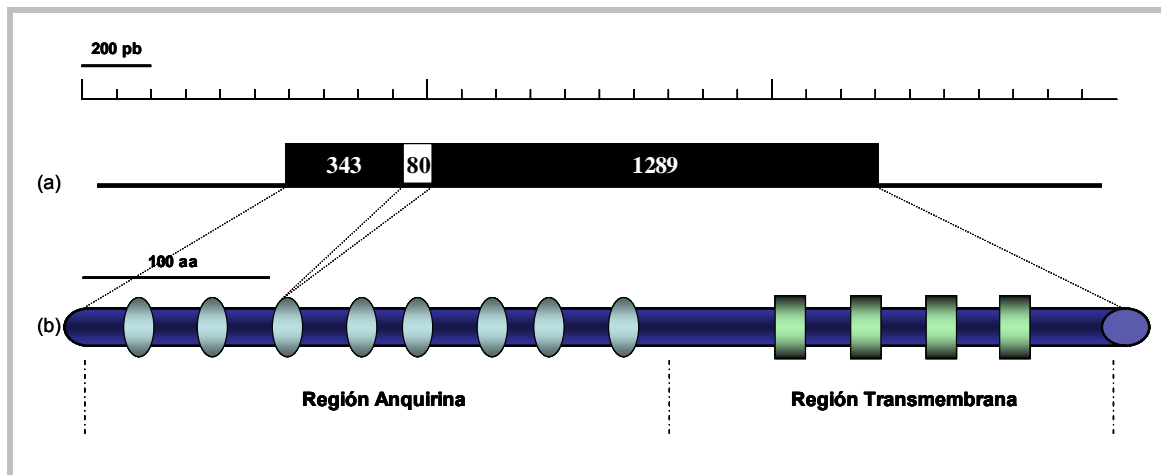


Figura 48.- Representación esquemática del DNA genómico (a) y la proteína codificada (b) por el gen *AtAnkTm28*, con las repeticiones anquirina (óvalos) y los dominios transmembrana (rectángulos).

### II.5.2 Patrón de expresión del gen *AtAnkTm28*

Como se señaló en el capítulo anterior, la amplificación de RT-PCR de este gen fue muy débil en silicua inmadura (Figura 36) y se decidió comprobar esta información. Se realizó un nueva RT-PCR utilizando cDNA de silicuas en etapas iniciales de desarrollo manteniendo las condiciones de PCR iniciales pero cambiando el número de ciclos de amplificación (pasando de 28 a 40 ciclos). Los resultados obtenidos en este experimento (Figura 49) permitieron confirmar lo observado en la primera experiencia.

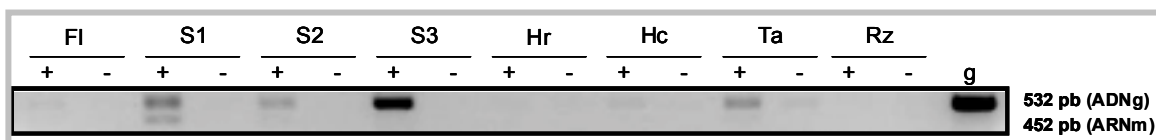


Figura 49.- Análisis por RT-PCR de la expresión del gen *AtAnkTm28* de la familia IV de genes *AtAnkTm* de *Arabidopsis thaliana*. Fl, flores; S1, silicuas inmaduras; S2, silicuas intermedias; S3, silicuas maduras; Hr, hoja de roseta; Hc, hoja caulinar; Ta, tallo; Rz, raíz. (+) representa el DNA codificante; (-) corresponde al control de contaminación de RNAm de cada órgano. g, control de DNA genómico. Los valores de amplificación de banda de DNA genómico y RNA mensajero (532 y 452 pares de bases respectivamente) son los esperados para el par de oligonucleótidos empleados.

El fragmento de cDNA amplificado en silicua inmadura (S1) fue clonado y secuenciado, confirmándose que correspondía al gen 28. El siguiente paso fue determinar en que parte de la silicua se expresaba: semillas o vaina. Para ello se colectó material vegetal de silicuas inmaduras de Columbia 0, en estadíos iniciales de desarrollo, separándose cuidadosamente las semillas de las vainas. Se extrajeron RNAs totales de éstos órganos y se utilizaron en análisis por RT-PCR. Previamente se determinó la calidad del cDNA utilizando como control el gen de *actina*. La amplificación de bandas permitió confirmar la calidad del cDNA en los cDNAs de vaina y de semilla (Figura 50). A continuación se realizó el mismo experimento con los iniciadores para el gen 28 y se encontró expresión específica del gen 28 en semilla (Figura 51). Para lograr amplificar la banda esperada se utilizó una reacción de PCR de 40 ciclos.

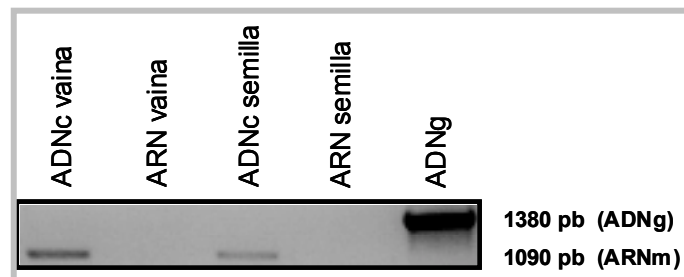


Figura 50.- Análisis de RT-PCR del gen de *actina* en vaina (cDNA) y semilla (cDNA) de silicuas inmaduras (S1) de *Arabidopsis thaliana*. Controles: RNA de vaina y semilla, y DNA genómico (DNAg).

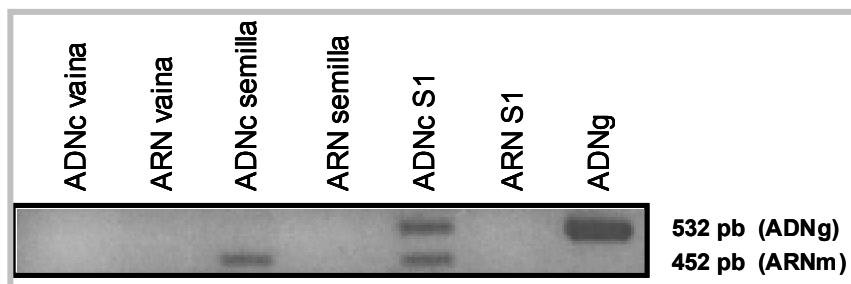


Figura 51.- Análisis de RT-PCR del gen *AtAnkTm28* en vaina (4 a 8 DDP), semilla inmadura (4 a 8 DDP) y silicua inmadura S1 (1 – 7 DDP). Para cada órgano se utilizó RNAm como control de contaminación. g, control de DNA genómico. Los valores de amplificación de banda de DNA genómico y RNA mensajero (532 y 452 pares de bases respectivamente) son los esperados para el par de oligonucleótidos empleados.

Finalmente, se analizó por northern blot la expresión del gen 28 utilizando una sonda diseñada en la región 3' UTR (*Untranslated region*). No fue posible detectar la expresión de este gen en ninguno de ellos por este método (Figura 52).

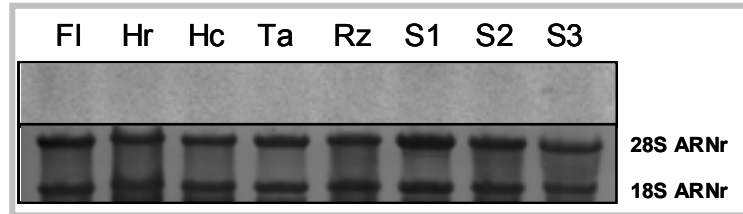


Figura 52.- Análisis por northern blot de la expresión del gen *AtAnkTm28* en diferentes órganos de *Arabidopsis thaliana*. FI, flores; S1, silicuas inmaduras; S2, silicuas intermedias; S3, silicuas maduras; Hr, hoja de roseta; Hc, hoja caulinar; Ta, tallo; Rz, raíz. 28S y 18S, RNA ribosomales.

### II.5.3 Patrón espacial de expresión del gen *AtAnkTm28*

Una vez comprobada la expresión específica del gen 28 en semilla en estadíos tempranos de desarrollo decidimos averiguar que parte de éste órgano era en donde se expresaba el gen para lo cual se realizaron hibridaciones *in situ*. Previamente, la calidad de los cortes se comprobó mediante hibridación con naranja de acridina (Figura 53). Este colorante se une al RNA presente en los órganos y por fluorescencia (tinción de color anaranjada) permite definir si el tejido está en óptimas condiciones.

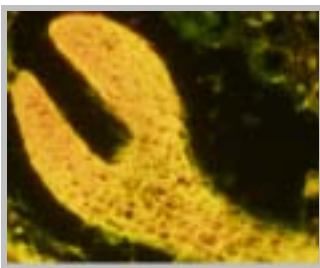


Figura 53.- Tinción con naranja de acridina en embrión de *Arabidopsis thaliana* en estado de torpedo.

La sonda antisentido para el gen 28 se une específicamente al embrión en estadio globular, al suspensor y al endospermo nuclear libre (Figura 54.A y C). En la hibridación con la sonda sentido no se puede apreciar tinción (Figura 54.B y D).

Se realizó un experimento adicional en forma paralela para verificar que las condiciones de hibridación *in situ* eran adecuadas y que los resultados



obtenidos no eran falsos positivos. Se utilizó una sonda del gen *abi3* (*Ab*scisic acid *i*nsensitive 3), cuya expresión es específica del desarrollo de embrión (gentileza de la Dra. Martine Devic, Universidad de Perpignan, Francia). La hibridación de la sonda antisentido del gen *abi3* se detectó específicamente en embrión de *Arabidopsis thaliana* (Figura 16.A), por lo tanto, las condiciones de hibridación empleadas fueron las correctas.

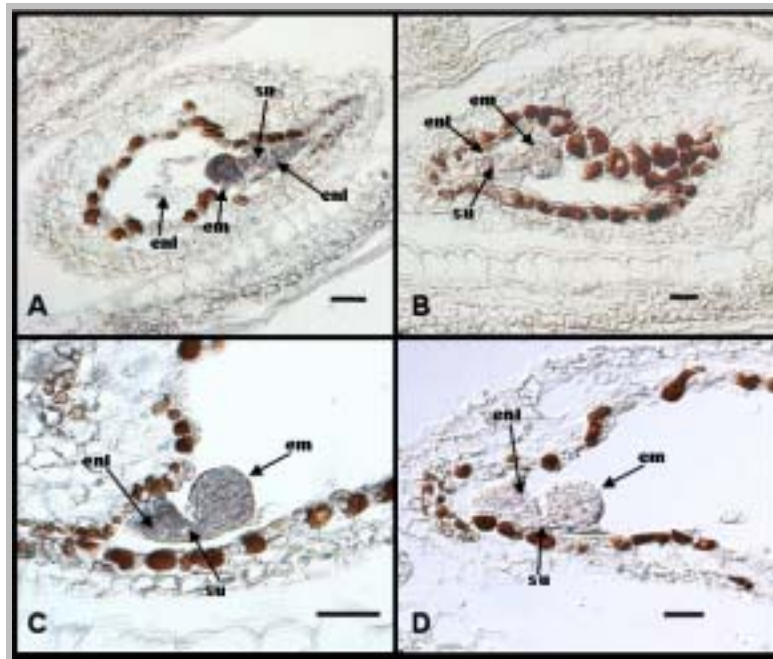


Figura 54.- Hibridación *in situ* del gen *AtAnkTm28* en silicuas inmaduras de *Arabidopsis thaliana*. (A) y (C) corresponden a la hibridación de la sonda antisentido, (B) y (D) a la hibridación de la sonda sentido. Abreviaturas: em, embrión; enl, endospermo nuclear libre; su, suspensor. Barras en A, B y D = 20  $\mu$ m; C = 50  $\mu$ m.

#### II.5.4 Localización subcelular de la proteína ATANKTM28.

La proteína codificada por el gen *AtAnkTm28* posee dominios transmembrana, por lo debería estar localizada en alguna de las membranas de la célula. Para averiguar en cual de ellas, se fusionó la región de los dominios transmembrana del gen 28 al gen informador *gfp*. Mediante bombardeo (biobalística) de células de epidermis de cebolla (*Allium cepa* L.) se determinó la localización del fragmento transmembrana de esta proteína (Figura 55). La fusión de la región transmembrana con la proteína GFP (TM28:GFP) mostró una localización en puntos de lo que posiblemente corresponda a la membrana citoplasmática de la célula de epidermis de cebolla



(Figura 55.B). Es probable que los puntos de mayor intensidad de fluorescencia correspondan a la localización de la proteína TM28:GFP en la misma membrana pero en planos distintos. Para este experimento se utilizó como control el vector de expresión sólo con GFP. La distribución típica de la GFP es difusa en el citoplasma y también se encuentra en el núcleo (Figura 55.C y D).

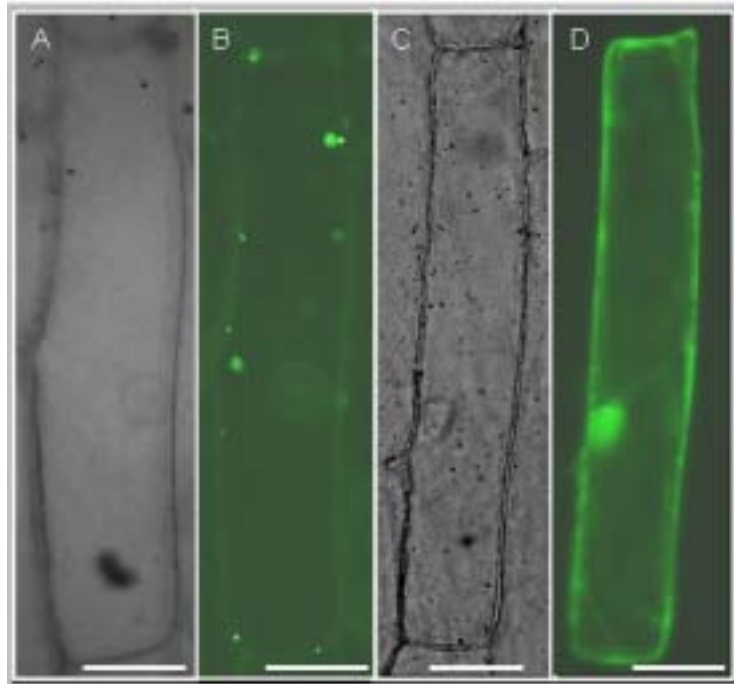


Figura 55.- Localización celular de la fusión de la región transmembrana codificada por el gen *AtAnkTm28* con la proteína *Green Fluorescent* (GFP) en células de catáfilo de cebolla. (A) y (C) células de catáfilo observadas por microscopio óptico correspondiente a la región transmembrana (TM28:GFP) y el control con GFP, respectivamente. (B) corresponde a TM28:GFP y (D) a GFP. Barras, 20  $\mu$ m.

## **DISCUSIÓN**

---

## **Capítulo I. Identificación de genes que se expresan específicamente durante el desarrollo temprano de la semilla de *Arabidopsis thaliana***

Las semillas son órganos complejos que combinan tejidos diploides procedentes del cigoto (embrión) y de la planta madre (paredes del ovario) y tejido triploide (endospermo). Durante la embriogénesis el cigoto se divide, adquiere el patrón morfológico propio del embrión, acumula reservas, se protege de la desecación, etc. (Willemsen and Scheres, 2004; Jürgens *et al.*, 1994; Hemerly *et al.*, 1999; Elster *et al.*, 2000). Al mismo tiempo, el pericarpo o cubierta de la semilla adquiere sus características protectoras (Haughn y Chaudhury, 2005) y el endospermo, en el caso de *Arabidopsis*, desaparece tempranamente (Olsen, 2004). Por tanto, el desarrollo de una semilla requiere de un funcionamiento de muchos programas genéticos diferentes que han de funcionar de manera coordinada espacial y temporalmente. Algunos de estos procesos son exclusivos de semilla por lo que es de suponer que han de existir genes cuya expresión tenga lugar exclusivamente en este órgano. Uno de los objetivos de esta tesis era la identificación de genes que intervienen en la embriogénesis en *Arabidopsis thaliana*. Para ello se han utilizado dos estrategias de análisis, una de ellas a nivel experimental y la otra, mediante búsqueda de información *in silico*.

### **I.1 Obtención de nuevas secuencias de ESTs**

Como se ha comentado, el número de ESTs de *Arabidopsis* depositados en las bases de secuencias puede ser limitante para los análisis *in silico*. Esto es especialmente cierto para algunos órganos. El número de ESTs disponibles de semilla inmadura en estadíos muy tempranos de desarrollo era muy bajo al inicio de esta tesis: 54 secuencias. Por ello se construyó una nueva genoteca de cDNA a partir RNAs extraídos de semillas inmaduras (genoteca ATISLA) y se secuenciaron 178 ESTs. A pesar de lo limitado del número de secuencias nuevas se obtuvieron ESTs correspondientes a 95 genes (47% de redundancia), y se incrementó en un 0,5% la presencia de ESTs de semilla inmadura en las bases de datos, y en un 330% la presencia de ESTs de

semillas en estadios tempranos de desarrollo. Para 2 de los 95 genes identificados no se habían obtenido ESTs hasta la fecha. Un 30,5% de los genes identificados en la genoteca ATISLA no poseen una función conocida. Destaca el hecho que 4 genes cuya función es desconocida (*At4g12960*, *At1g08480*, *At3g08610* y *At4g00585*) están entre los que poseen una mayor cantidad de ESTs en las secuenciadas de ATISLA. Resultaría muy interesante poder estudiar y analizar más en profundidad cada uno de ellos, en especial el gen *At4g12960* que tiene 8 ESTs en la genoteca ATISLA (4,5%) y 118 en las otras genotecas.

Al comparar los resultados de la genoteca ATISLA (2 a 6 días después de antesis) con los resultados obtenidos por White y colaboradores (2000) (5 a 13 días después de floración; Cat# 5564) (Cuadro 17) es posible encontrar algunas diferencias en cuanto a las categorías obtenidas que representan cambios en la expresión génica a lo largo del desarrollo de la semilla. Destacaría que la genoteca ATISLA presenta, respecto a la genoteca 5564:

- Muchos menos, de hecho, ningún EST de la categoría “Reserva de nutrientes”. La acumulación de reservas comienza en estadios intermedios de desarrollo por lo que es lógico que no aparezcan estos genes en ATISLA (White *et al.*, 2000).
- Mucho mayor representación de genes relacionados con el ciclo celular (9 veces más). La mayor parte de las divisiones que tienen lugar en la semilla ocurren durante las fases iniciales de su desarrollo, por lo que es lógico que haya mayor expresión de este tipo de genes en ATISLA (Jürgens *et al.*, 1991).
- Mayor representación de genes de defensa (5 veces más). Muchos genes de defensa se expresan en las semillas. Por ejemplo, los genes más abundantes que se expresan específicamente en la capa de transferencia del endospermo de maíz codifican proteínas antipatogénicas, señalando un papel de estas células en la protección del endospermo contra la entrada de patógenos (Thompson *et al.*, 2001).
- 3,5 veces más representación de genes relacionados con la traducción.

En el resto de categorías las diferencias no son tan importantes (diferencias menores a 2,5 x). Entre las categorías funcionales que poseen mayor similitud destacan las relacionadas con el metabolismo, tanto de carbohidratos, como de lípidos, secundario y fotosíntesis. También es semejante la contribución de los genes de función desconocida, algo menor en ATISLA, pero es preciso tener en cuenta que el análisis de White y colaboradores fue realizado hace 5 años, por lo que el aumento de disponibilidad de información respecto a la función de algunos genes podría variar esta relación respecto al porcentaje inicial.

Cuadro 17.– Categorías funcionales en genotecas de ESTs de semilla inmadura

Categoría funcional	ATISLA <sup>1</sup>	5564 <sup>1</sup>
	(2 - 6 DDA <sup>2</sup> )	(5 - 12 DDA)
Ciclo de división celular	2,8 <sup>3</sup>	0,3
Defensa	2,8	0,6
Desarrollo	5,1	3,0
Detoxificación de oxígeno	4,5	1,7
Energía	2,2	1,1
Fotosíntesis	1,7	1,5
Metabolismo de ácidos nucleicos	0,6	1,6
Metabolismo de aminoácidos	1,1	2,6
Metabolismo de carbohidratos	9,0	6,7
Metabolismo de lípidos	4,5	4,7
Metabolismo secundario	2,2	2,3
Procesamiento de proteínas	6,7	3,9
Regulación de la transcripción	3,4	5,6
Reserva de nutrientes	0,0	14,4
Traducción	13,5	3,9
Tráfico subcelular y transporte	6,2	2,3
Transcripción, splicing	1,1	1,5
Desconocida	32,6	42,4

1: Código de las genotecas.

2: DDA: Días Después de Antesis.

3: Datos expresados en porcentaje de ESTs.

## 1.2 Identificación de genes de expresión específica en semilla inmadura

La era de la genómica ha proporcionado a los investigadores una enorme cantidad de información que está revolucionando los estudios moleculares de la vida. La secuenciación de genomas enteros ha permitido identificar todos los posibles genes codificados. Sin embargo, la simple secuencia de un gen proporciona datos limitados sobre su función. Un primer paso para la determinación de la función de un gen desconocido es identificar

dónde, cuándo y en respuesta a qué se transcribe. En este sentido, la acumulación de secuencias de ESTs y de datos de hibridaciones de micromatrices proporciona una nueva oportunidad al investigador para el estudio inicial de genes de función desconocida.

El análisis *in silico* o northern virtual consiste en el estudio de los patrones de expresión de uno o más genes basándose en la abundancia de secuencias en las distintas genotecas de cDNA estudiadas. Estas secuencias pueden ser ESTs o secuencias cortas de cDNA obtenidas mediante la técnica de SAGE (*Serial Analysis of Gene Expression*). Una de las principales ventajas del northern virtual es la rapidez y el bajo coste una vez se han generado las secuencias (Federova *et al.*, 2002; Bernstein *et al.* 1996; Welle *et al.* 1999; Bortoluzzi *et al.* 2000). Este sistema ha sido utilizado con éxito en humanos, especie para la cual hay depositadas más de 7 millones de ESTs en las bases. Por ejemplo, basándose únicamente en la abundancia de las ESTs, Vasmatzis *et al.* (1998) identificaron *in silico* tres genes que se expresan exclusivamente en próstata, Itoh *et al.* (1998) identificaron diez genes de expresión predominante en granulocitos, y Miner y Rajkovic (2003) un centenar genes de expresión predominante en placenta. La combinación del análisis de abundancia de ESTs con SAGE permitió la identificación de genes de expresión específica en endotelio (Huminiacki y Bicknell, 2000). Existen para genes humanos servidores en red que permiten un fácil y sencillo análisis de la gran cantidad de información acumulada en las bases de secuencias, como HsAnalyst para ESTs (Baranova *et al.*, 2001) y el Xprofiler para SAGE (Chen *et al.*, 2003).

Para plantas la cantidad de información disponible es mucho menor que para humanos. El número de ESTs de *Arabidopsis* depositados es unas 15 veces menor ([http://www.ncbi.nlm.nih.gov/dbEST/dbEST\\_summary.html](http://www.ncbi.nlm.nih.gov/dbEST/dbEST_summary.html); Rudd, 2003) y los datos de SAGE son mucho menos abundantes y de momento limitados a unas pocas condiciones u órganos (Robinson *et al.*, 2004; Fizames *et al.*, 2004; Ekman *et al.*, 2003; Jung *et al.*, 2003; Lee y Lee, 2003). Ante esta limitación en los datos podemos pensar en la complementación con datos de

análisis experimentales públicamente accesibles, concretamente, las bases de datos de hibridaciones de micromatrices.

Uno de los objetivos de este trabajo era la identificación de genes de expresión específica durante el desarrollo temprano de la semilla de *Arabidopsis*. Para ello se llevó a cabo un análisis *in silico* de genes específicos en el desarrollo de la semilla empleando bases de secuencias de ESTs y de datos de hibridaciones de micromatrices (Meta-Analyzer, Genevestigator<sup>®</sup>; <https://www.genevestigator.ethz.ch/>). Siguiendo los criterios establecidos tanto para la selección por ESTs en primer lugar y por micromatrices en el segundo, fue posible la identificación de 49 genes que se expresan mayoritariamente en semilla inmadura. Dentro de este conjunto se encontraron algunos genes cuya expresión ha sido descrita anteriormente como específica del desarrollo de la semilla, como son *abi3* (*Abscisic acid insensitive 3*; Giraudat *et al.*, 1992); *At1g48130*, que codifica una peroxirredoxina (PER1; Haslekas *et al.*, 1998); *At1g67100*, que es homólogo al gen *Bn15D17A* de embrión de *Brassica napus* (Colza) (Dong *et al.*, 2004); *At5g07190* y *At5g55240*, que codifican proteínas específicas de embrión (Nuccio y Thomas, 1999). Los datos de hibridación de micromatrices obtenidos a partir del servidor GEV demuestran que, aunque en algunos casos existe expresión baja en otros órganos, la expresión es mayoritariamente específica en semillas inmaduras (Figura 56). Todo esto junto con los resultados experimentales obtenidos por RT-PCR semicuantitativa e hibridación *in situ* demuestran que, pese a las limitaciones, ya es posible utilizar métodos *in silico* de estudio de la expresión génica en *Arabidopsis*.

El número de genes identificados (49) puede ser considerado bajo y puede ser consecuencia de un número insuficiente de ESTs en las bases de datos. Rudd (2003) mostró que sólo 16.115 genes de *Arabidopsis* (aproximadamente el 60%) están representados en las bases de datos de ESTs. Aunque parcialmente resuelto gracias al trabajo realizado durante esta tesis, este problema es especialmente grave en el caso de los genes que se expresan en semilla inmadura temprana. Es probable que no se hayan identificado muchos genes cuya expresión sea muy baja o muy localizada. Por ejemplo, genes que se expresen exclusivamente y en nivel medio o bajo en el embrión en estadios muy tempranos del desarrollo (globular a corazón) es

difícil que aparezcan en las bases de ESTs, y no habrán sido seleccionados. Uno de estos casos es el gen *At1g07710* (*AtAnkTm28*, ver capítulo II de resultados). Este gen se expresa de manera específica y exclusiva en el embrión al inicio de su desarrollo y sin embargo no aparece en la lista de genes seleccionados porque no hay ESTs correspondientes en las bases de secuencias. Un problema adicional es que en las micromatrices de Affymetrix no están representados todos los posibles genes identificados en el genoma de *Arabidopsis*. Un 7,5 % de los genes seleccionados por ESTs no aparecían en la matriz y no pudieron ser analizados.

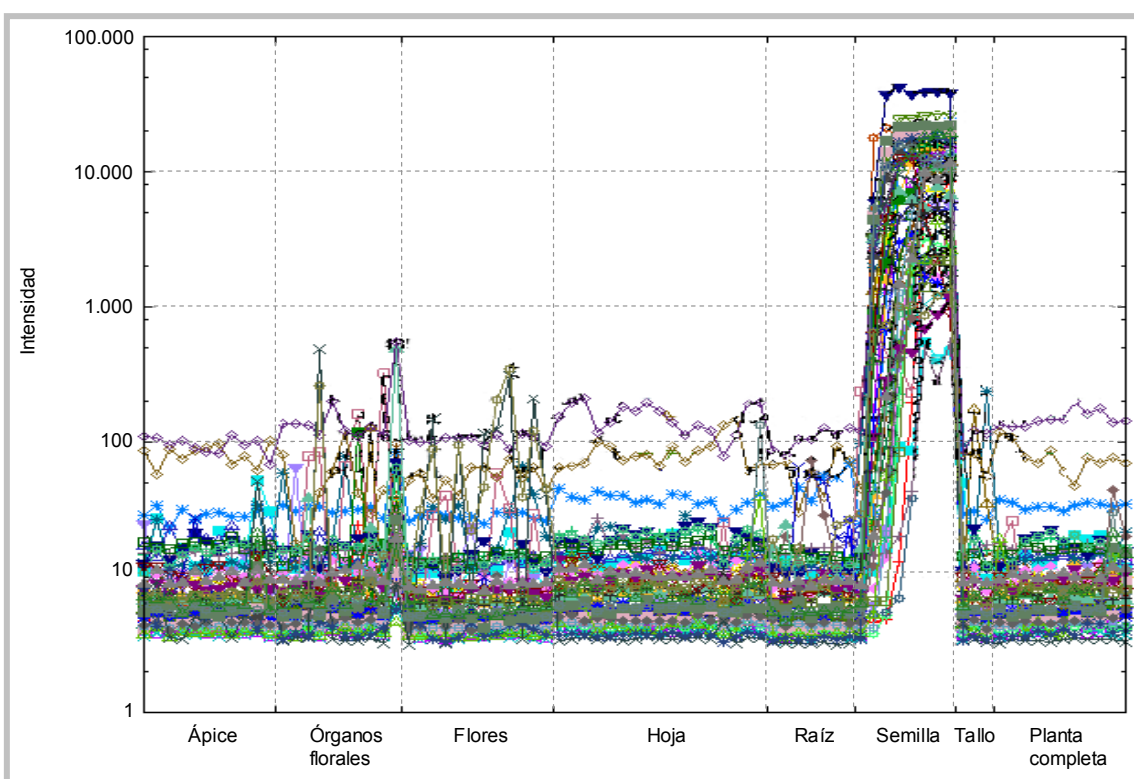


Figura 56.- Patrones de expresión de los 49 genes seleccionados mediante análisis *in silico* basados en datos de micromatrices. Eje X, diferentes órganos vegetales. Eje Y, valores de intensidad de hibridación. Cada punto de la gráfica corresponde a la media de tres intensidades de hibridación. Datos obtenidos a partir de AtGenExpress, *Gene Expression Visualization* (GEV) (<http://www.weigelworld.org/resources/microarray/AtGenExpress/>) (Schmid *et al.*, 2005).

Además de los problemas técnicos señalados, cabría preguntarse si el bajo número de genes seleccionados no responde a un hecho real. Por ejemplo, podría sorprender el reducido número de genes seleccionados que codifican factores de transcripción. Sin embargo, recientes análisis globales de datos de expresión indican que el número de genes de factores de transcripción que se expresan específicamente durante el desarrollo de la



semilla es relativamente bajo comparado con otros órganos (Lehti-Shiu *et al.*, 2005; Ma *et al.*, 2005). Se analizó la expresión de diversos genes MADS-box en diferentes tejidos de *Arabidopsis* y se encontró que muchos se expresan en cultivo de tejidos embrionarios, pero pocos de ellos se expresan exclusivamente en este tejido (Lehti-Shiu *et al.*, 2005). De forma similar, el número de genes que codifican factores de transcripción que se expresan específicamente en silicuas en desarrollo es relativamente bajo, comparado con otros tejidos (Ma *et al.*, 2005). Se ha propuesto que muchos de los procesos de desarrollo que ocurren durante la embriogénesis son activos también durante algunos procesos del desarrollo vegetativo de la planta, por ejemplo, en los meristemas, o durante el desarrollo floral, y esto hace que el número de genes de expresión exclusiva durante la embriogénesis sea bajo (Lehti-Shiu *et al.*, 2005). Existen ejemplos en la literatura de genes que codifican proteínas necesarias para el desarrollo del embrión que también se expresan en otros órganos de la planta. Por ejemplo, el gen *FAC1* (*Embryonic Factor 1*) que se expresa en la semilla al inicio de su desarrollo y cuya mutación produce letalidad en el embrión (Xu *et al.*, 2005) no aparece en la lista porque también se expresa durante el desarrollo vegetativo. De la misma manera, el gen *AtML1* (*Arabidopsis thaliana meristem L1 layer*) que interviene en el establecimiento del patrón apical-basal del embrión, también se expresa en el meristema apical vegetativo y en los botones florales (Lu *et al.*, 1996). Un análisis más sistemático se llevó a cabo partiendo de una muestra tomada al azar de 68 genes cuya mutación produce letalidad en el embrión, en concreto, todos los genes que producen mutaciones letales en embrión presentes en la base [www.seedgenes.org](http://www.seedgenes.org) y que se encuentran en el cromosoma 1. Ninguno de estos genes está entre los 49 seleccionados. Determinamos las razones por las que no fueron escogidos:

- Para 9 genes no se ha secuenciado ninguna ESTs (13 %) y para otros 4 (6 %) solo se han secuenciado ESTs a partir de genotecas de cDNA no informativas (provenientes de mezclas de órganos).
- Para 54 de los 68 genes (79 %) se han secuenciado ESTs en genotecas construidas a partir de órganos que no son semilla inmadura. En esta misma situación se encuentran otros genes cuyas mutaciones

producen alteraciones bien caracterizadas en el desarrollo del embrión como *pin-formed4*, *fackel*, *prolifera* y *globular arrest1*. Hay que tener en cuenta que hemos sustraído de la lista aquellos genes que presentan ESTs en genotecas de semilla seca.

- Un gen superó los criterios de selección por ESTs, pero no el de micromatrices (gen *acc1*, que codifica una acetil-CoA carboxilasa). En la misma situación se encuentran, por ejemplo, *fusca3* y *PEI1*.

Se concluye, por un lado, que una alta proporción de genes implicados en la formación del embrión no son específicos de semilla, y por otro lado, que la existencia de un mayor número de ESTs de semilla inmadura podría haber incrementado el número de genes seleccionados. Otra posibilidad habría sido el uso de criterios de selección menos restrictivos en la selección por micromatrices, pero esto, probablemente, hubiera producido unos resultados menos específicos.

El análisis de las categorías funcionales de los genes representados en la lista seleccionada nos presenta una imagen de los procesos más específicos del desarrollo de la semilla. En este sentido, no es casualidad que la categoría más representada sea de reserva de nutrientes, ya que son genes que, por un lado se expresan en semilla a niveles muy altos y, por otro, tienen una función exclusiva en la semilla. No es pues de extrañar que la proporción de genes en la lista final sea mucho mayor que para el total del genoma (Cuadro 6). El porcentaje de genes relacionados con el estrés abiótico se incrementa con la selección y ocurre lo mismo con los genes relacionados con el metabolismo de carbohidratos, ya que las semillas sufren un proceso de desecación y acumulan carbohidratos como sustancias de reserva.

Es destacable la abundancia de genes que codifican proteínas relacionadas con la pared celular: un gen que codifica la xiloglucan:xiloglucosil transferasa, implicada en la síntesis de pared, un gen que codifica una extensina y otro gen que codifica una pectinesterasa. Ésta es una indicación del alto índice de síntesis de paredes celulares nuevas durante el desarrollo de la semilla, y podría también ser una indicación de la importancia de los componentes específicos de la pared en la señalización célula-célula para la

coordinación de los programas génicos durante el desarrollo del embrión (Souter y Lindsey, 2000), un efecto observado en embriones inmaduros del maíz (Jose-Estanyol *et al*, 1992). La presencia de los genes implicados en el desarrollo no es perceptiblemente más alta en los genes seleccionados comparados con el genoma entero (6.0 %). Esto no es sorprendente debido a que el genoma entero contiene numerosos genes implicados, por ejemplo, en el desarrollo de la flor o de la raíz. Otro gen posiblemente involucrado en el desarrollo del embrión sea *At1g48660*, que codifica una proteína de la familia GH3, y que responde a auxinas. El transporte polar de auxinas es básico para el establecimiento del eje de apical-basal del embrión (Bai *et al.*, 2000).

Dentro de los genes seleccionados y de cara a futuros proyectos, resultan de especial interés los genes de función desconocida identificados cuyo máximo nivel de expresión tiene lugar en etapas iniciales o intermedias del desarrollo de la semilla. Entre ellos destacan:

- *At1g62060*: codifica una proteína de 151 aa de función desconocida con un pI (punto isoeléctrico) de 11,1. Se han secuenciado 58 ESTs correspondientes a este gen. No presenta ningún motivo o señal identificables.
- *At1g67100*: codifica una proteína de 234 aa, de función desconocida con un pI de 8,1. Se han secuenciado 6 ESTs correspondientes a este gen. Presenta una región o dominio de unos 100 aa presente en varias proteínas de plantas pero del cual se desconoce la función (código PFAM número DUF260). Es homólogo al gen *Bn15D17A* de *Brassica napus* cuya expresión es específica de semilla (Dong *et al.*, 2004).
- *At1g80090*: codifica una proteína de 403 aa de función desconocida con un pI de 7,4. Existen 5 ESTs correspondientes a este gen en los bancos de secuencias y la proteína que codifica contiene dos dominios CBS (Cistatión betasintasa). Estos dominios normalmente se encuentran en número de 2 ó 4 y se encuentran en muchas proteínas de diversas funciones. Pueden actuar como lugares de unión a derivados de la adenosina, pueden regular la función de enzimas, actuar como sensores

del estado de la energía celular al ser activados por AMP e inhibidos por ATP, también pueden intervenir en el tráfico intracelular o en interacciones proteína-proteína (Bateman, 1997; Ponting, 1997; Sintchak, 1996).

En conclusión, a pesar de los problemas técnicos asociados a la reducida cantidad de ESTs disponibles de *Arabidopsis*, se ha demostrado que el uso de la sustracción *in silico* es posible y genera los resultados esperados (Becerra *et al.*, 2006). El mismo método podría ser potencialmente aplicable a otras especies dependiendo de la disponibilidad de ESTs obtenidos a partir de órganos aislados. El principal problema es la disponibilidad de datos de micromatrices, que es mucho más limitado en otras especies y, de momento, no existen aplicaciones para su análisis masivo.

## **Capítulo II. Genes que codifican proteínas con repeticiones anquirina y dominios transmembrana implicados en la embriogénesis de *Arabidopsis thaliana*.**

### II.1.- Repeticiones anquirina en Arabidopsis

Se han identificado un total de 509 repeticiones anquirina (ANK) en 105 proteínas de Arabidopsis, lo cual representa que un 0.4% de los genes de Arabidopsis codifican proteínas con estas repeticiones. Este número es más alto que el estimado anteriormente de 0.25% (Jebanathirajah *et al.*, 2002), pero es similar a los porcentajes estimados para los seres humanos, *C. elegans* y *Drosophila*. Pocos de los aminoácidos en las repeticiones ANK están bien conservados (Cuadro 6). El uso de criterios más exactos para identificar repeticiones ANK ha permitido reconocer muchas repeticiones previamente no anotadas. Al evaluar la conservación de las secuencias de las repeticiones resulta obvio que las repeticiones situadas en los extremos de las agrupaciones en tándem se desvían más del consenso general que las localizadas centralmente, cosa que fue observada también en las proteínas animales (Bork, 1993).

Aunque pocos de los 33 aminoácidos que componen las repeticiones ANK se conservan totalmente, existen algunas posiciones siempre ocupadas por aminoácidos hidrofóbicos. Esto ocurre tanto en Arabidopsis como en animales. Se sabe a partir de datos de proteínas animales que esta conservación es necesaria para mantener la estructura secundaria y esto es esencial para cumplir la función en las interacciones de la proteína-proteína (Figura 57) (Bork, 1993; Rohde y Bork, 1993). Aunque no se ha cristalizado ninguna proteína de plantas con repeticiones ANK, y, por tanto, no se conoce su estructura tridimensional, la conservación de la secuencia consenso y su presencia en agrupaciones en tándem, como sucede en animales, sugiere que la estructura y función en plantas han de ser similares.

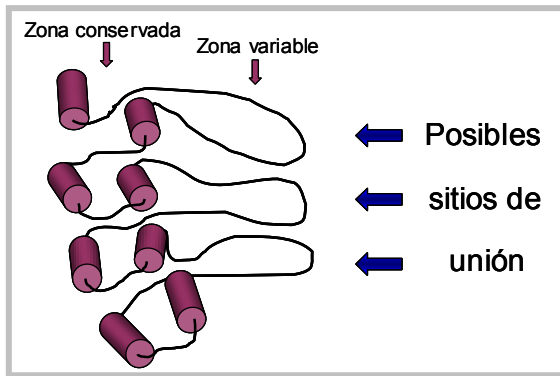


Figura 57.- Modelo esquemático de la estructura de las repeticiones anquirina. Los cilindros granates representan las hélices  $\alpha$  de cada una de las repeticiones.

Muchas proteínas de plantas que contienen repeticiones ANK son proteínas multidominio en las cuales las repeticiones ANK se combinan con otros módulos estructurales (Cuadro 8). La presencia de las repeticiones ANK en proteínas de plantas tan diversas sugiere que realizan una función general y no una actividad enzimática concreta. En animales, las repeticiones ANK se encuentran también en proteínas de muy diversa localización celular y función. Contienen repeticiones ANK por ejemplo: subunidades de factores de transcripción, reguladores de sistemas transcripcionales, proteínas intrínsecas de membrana que regulan la diferenciación de tejidos, un determinante sexual de nematodos, proteínas reguladoras de fosfolipasas, toxinas de arácnidos, intercambiadores iónicos  $\text{Cl}/\text{HCO}_3$ ,  $\text{Na}/\text{Ca}$ , la ATPasa  $\text{Na}/\text{K}$ , el receptor  $\text{IP}_3$ , rianodina, canales de  $\text{Na}$  dependientes de voltaje o proteínas de adhesión celular (Givskov *et al.*, 1988; Larsson *et al.*, 1998; Kiyatkin *et al.*, 1993; Michaely *et al.*, 2002, Chang y Low, 2003; Bennett y Baines, 2001).

Pocas proteínas que contienen repeticiones ANK han sido caracterizadas en vegetales (Becerra *et al.*, 2004). Entre sus funciones se pueden citar:

- Defensa frente a patógenos: ANK1 de tabaco, AKR2, ART2, NPR1, ACD6 de *Arabidopsis*;
- Factores de transcripción: CAMTAs;
- Canales de potasio: Familia AKT de *Arabidopsis*, SKT1 en patata;
- Quinasas: APKs en *Medicago*;
- Diferenciación celular: AKR de *Arabidopsis*;
- Proteínas de unión: CAO y ACBP2;
- Desarrollo embrionario: EMB506 de *Arabidopsis*.

Aunque no se ha demostrado experimentalmente, estos datos apoyan la idea de que en plantas, como en animales, las repeticiones ANK no tienen una función específica sino general. Dado que la secuencia consenso de las repeticiones de animales y plantas es semejante, la función probablemente sea la misma e intervengan en interacciones proteína-proteína.

## II.2 Genes *AtAnkTm* de *Arabidopsis thaliana*

El grupo más abundante de proteínas estructuralmente similares que contienen repeticiones ANK cuenta con 40 elementos: 37 proteínas que poseen repeticiones ANK y dominios transmembrana, y tres más que pueden representar formas truncadas ya que contienen sólo repeticiones ANK. Los 40 genes *AtAnkTm* están divididos en seis familias. La evolución de estas familias de genes ha sido compleja. Los genes *AtAnkTm* están distribuidos por todo el genoma pero no uniformemente. Este tipo de distribución parece ser común de otras familias de genes y, por ejemplo, es similar a la distribución de los genes *bHLH*, aunque las regiones de alta y baja densidad no son las mismas en ambos casos (Toledo-Ortíz *et al.*, 2003). Las duplicaciones inter- e intracromosómicas del genoma y las duplicaciones en tándem de los grupos de genes han jugado un papel en amplificar el número de genes *AtAnkTm*. 13 de estos genes (33%) están duplicados en el genoma. Las duplicaciones en tándem también han jugado un papel importante en la evolución de estos genes. 16 de los genes *AtAnkTm* (40%) están localizados en grupos organizados en tándem. En total, 22 genes *AtAnkTm* (55%) están duplicados en el genoma de alguna manera. Estos grupos de genes en tándem son muy comunes en *Arabidopsis*. Haberer y colaboradores (2004) encontraron que el 17 % de los genes de *Arabidopsis* se encuentran en tándem o en segmentos duplicados. Este valor es mucho menor respecto al observado para los genes *AtAnkTm*, es decir, que estos genes están especialmente repetidos en el genoma. Esta mayor repetición podría haber permitido una diversificación de funciones que, por ejemplo, se reflejaría en una variedad de patrones de expresión dentro de los genes de una misma familia.

### II.3 Patrones de expresión de los genes *AtAnkTm* de *Arabidopsis thaliana*.

Se han empleado tres tipos de técnicas en el estudio de los patrones de expresión, una de carácter experimental y dos de ellas basadas en la consulta de bases de datos. Los tres métodos contribuyen al conocimiento de los patrones de expresión de cada gen, con sus ventajas e inconvenientes. En general, los resultados son coincidentes, aunque no siempre. La falta de coincidencia en algunos casos puede ser debida al diverso origen de las muestras empleadas en cada caso. Esto es especialmente cierto en el caso de las inducciones por diferentes tratamientos ya que en estos casos los tiempos y las condiciones concretas del tratamiento pueden variar de uno a otro experimento. El estado fenológico del material biológico empleado en cada uno de los ensayos también puede ser diferente. Otro motivo de las divergencias puede ser el limitado número de ESTs disponibles para algunos órganos de *Arabidopsis*, en especial para ciertos tipos de estrés, que hacen que no estén tan representados, o no lo estén en absoluto. En cuanto a los resultados de hibridaciones de micromatrices podemos apreciar dos problemas adicionales, la falta de información para algunos de los genes y la posibilidad de hibridaciones cruzadas entre genes cuyas secuencias son muy similares. Por último, el uso de RT-PCR semicuantitativa proporciona datos cualitativos interesantes pero, obviamente, el uso de otras técnicas como la *Real Time* PCR pueden aportar datos cuantitativamente más exactos aunque a costa de una mayor cantidad de trabajo y un costo mayor.

El número de secuencias depositadas en las bases de datos de ESTs de los diferentes tipos de genotecas es variable para los genes de las diferentes familias *AtAnkTm*. Hasta la fecha se han depositado 196 secuencias de EST correspondientes a 23 genes *AtAnkTm*. El 25% corresponden al gen 11 (gen *acd6*). Para diez de los genes que poseen ESTs se han depositado un número mayor o igual a cinco. No existe una asociación entre las ESTs de los genes *AtAnkTm* y algún tipo de genoteca en particular, encontrándose genes cuyas ESTs provienen de diferentes órganos de la planta, estados fenológicos y condiciones de crecimiento y desarrollo normales o por tratamientos de estrés.



Los resultados del patrón de expresión por RT-PCR de los diferentes órganos y condiciones de estrés de algunos de los genes *AtAnkTm* indican que algunos de ellos se expresan en muy diversos órganos, sugiriendo que pueden tener funciones generales o pleiotrópicas (25, 29, 32), mientras que otros tienen una expresión más restringida y quizás funciones más específicas. Uno de los niveles más altos de amplificación que se han detectado por RT-PCR corresponde al gen 11 (gen *acd6*), concretamente en hojas. En algunos casos fue posible encontrar expresión sólo bajo condiciones de estrés (3, 9 y 13). En otros casos la expresión es específica de ciertos órganos como el gen 28 en semilla inmadura, el gen 11 en hoja o los genes 14 y 17 en raíz.

El análisis de micromatrices mediante *Gene Expression Visualization* (GEV) permite observar gráficamente el patrón de expresión del conjunto de genes de una familia *AtAnkTm* en los diferentes órganos de la planta. En términos generales, para el conjunto de genes *AtAnkTm* de los que se tiene información (29 genes), se encontró que una intensidad de hibridación alta observada en GEV se correspondía con un mayor número de ESTs depositadas en las bases de datos y a su vez, estos datos de expresión eran confirmados con los resultados de RT-PCR de los órganos. Por ejemplo, el gen 11 posee una de las más altas intensidades de expresión por GEV, especialmente en hojas, lo que coincide con lo encontrado por RT-PCR y al analizar la búsqueda de las bases de datos de ESTs. Los resultados previos de Lu y colaboradores (2003) son confirmados con estos resultados, quienes encontraron expresión del gen *acd6* (11) en hojas mediante análisis northern. La herramienta GEV es muy útil si se desea tener una aproximación respecto al posible patrón de expresión de un gen de *Arabidopsis thaliana*. Existe una completa información en las bases de datos de micromatrices para la mayoría de los genes, sin embargo, hasta la fecha no se encuentra disponible para la totalidad de los genes de esta especie.

Como valoración global de los resultados obtenidos vemos que los patrones de expresión varían mucho de unas familias a otras, e incluso dentro de una misma familia. Aunque no se poseen datos completos para todos los genes resulta probable que algunos de los genes descritos en este trabajo no

sean funcionales y, por tanto, estrictamente no puedan ser definidos como genes sino como pseudogenes. Por ejemplo, para el gen 6, que forma parte del tándem de siete genes de la familia I en el cromosoma IV, no se ha detectado expresión por RT-PCR ni por datos de hibridación de micromatrices, ni hay ESTs en las bases de secuencias. La falta de funcionalidad de algunos de los genes dentro de repeticiones en tándem no sería extraña, aunque, desde luego, el hecho de que no se haya detectado expresión no descarta que no pueda tener una expresión muy localizada o en respuesta a unas condiciones no estudiadas aquí. Otros ejemplos de genes repetidos con diferencias en la expresión son el gen 12, que se transcribe en flor y silicua, y el gen 13. En el caso de los genes 28, 33 y 34, repetidos en diferentes localizaciones del genoma, mientras que el gen 33 se expresa de manera más o menos constitutiva, el gen 28 hemos visto que tiene una expresión específica en semilla inmadura y el gen 34 parece expresarse únicamente en polen. Curiosamente, los tres genes truncados (sin dominios transmembrana) parecen ser activos transcripcionalmente.

Observamos, por tanto, una amplia variedad de patrones de expresión, desde genes de expresión en casi todos los órganos hasta genes de expresión restringida a ciertos órganos o en respuesta a ciertas condiciones o tratamientos. Esta variabilidad se observa incluso entre genes de la misma familia, e, incluso, entre genes duplicados. Cuando se comparan las familias *AtAnkTm*, las secuencias de las repeticiones ANK no están bien conservadas entre familias pero si lo están dentro de la misma familia. Dado que la especificidad de interacción de las repeticiones ANK parece estar determinada por su secuencia de aminoácidos (Bennett, 1992), esta conservación dentro de las familias sugiere que las repeticiones ANK de la misma familia pueden interactuar con las mismas o similares proteínas. Por otro lado, la gran variabilidad de patrones de expresión, incluso dentro de una misma familia, sugiere que aunque las interacciones de cada miembro de una misma familia puedan ser con proteínas similares, quizás cada proteína interacciona en unas condiciones u órganos diferentes, cumpliendo funciones más o menos específicas, según el caso.

#### II.4 Análisis de líneas mutantes de inserción de T-DNA

Se analizaron numerosas líneas mutantes con inserción de T-DNA para bastantes de los genes *AtAnkTm* de Arabidopsis y en solamente dos de ellas fue posible detectar diferencias respecto a la línea silvestre a nivel de fenotipo, concretamente para los genes 2 y 9, y en ambos casos las mutaciones producen letalidad en el grano de polen, si bien estos datos deberían ser confirmados con un segundo alelo de la mutación. Para otro de los genes, el gen 11 (*acd6*), ya se había descrito previamente la mutación *accelerated cell death 6*, que se caracteriza por una inducción de la muerte celular programada por respuesta hipersensible en ausencia de patógeno (Rate *et al.*, 1999).

En general, el análisis de un gran número de líneas mutantes de los genes *AtAnkTm*, la obtención de un escaso número de líneas con un fenotipo distinguible del silvestre y la similitud existente entre los genes *AtAnkTm*, indican que es probable que si se bloquea la función de uno de estos genes, ésta sea asumida por otro de los genes de la misma familia. Estas conclusiones también son indicadas por Lu y colaboradores (2003), quienes señalaron que la no obtención de mutantes con pérdida de función en la familia de genes tipo *acd6* (genes *AtAnkTm*) posiblemente sea debida a funciones sobrepuestas o redundantes de los genes.

La letalidad en el polen puede significar que los genes responsables juegan un papel básico en algún aspecto de la gametogénesis o bien, son necesarios para alguna función celular básica. Los datos de expresión indican que el gen *AtAnkTm9* se transcribe en muchos de los tejidos por lo que es de suponer que cumpliría una función celular básica cuya carencia produce letalidad en el gameto masculino aunque no en el femenino, quizás porque en este caso la planta madre es capaz de suplir la función perdida por la mutación.

El gen *AtAnkTm2* tiene un patrón de expresión mucho mas restringido y se transcribe en grano de polen y en semilla en etapas intermedias de desarrollo. Su mutación produce letalidad en el grano de polen a la vez que cambios en la estructura de la superficie (exina). Aunque hay numerosos trabajos que describen la estructura de la superficie del grano de polen en

muchas especies (Scott, 1994) apenas se conocen los factores genéticos implicados en este proceso. Se sabe que muchos orgánulos celulares intervienen en la formación de las estructuras de la pared: microtúbulos, matriz extracelular, retículo endoplasmático, membrana plasmática, etc. (Paxson-Sowers *et al.*, 2001). Algunos estudios sugieren que el reticulado de la exina se debe a invaginaciones de la membrana citoplasmática (Takahashi y Skvarla, 1991). Los datos disponibles para el gen 2 indican que podría estar implicado en la formación de la exina. La proteína ATANKTM2 posee dominios transmembrana, por lo que podría estar localizada en la membrana citoplasmática. Como veremos a continuación, algunas de las funciones propuestas para este tipo de proteínas pueden estar relacionadas con el transporte de sustancias a través de la membrana, lo cual sugiere que la proteína ATANKTM2 podría estar implicada en el transporte de productos hacia la pared en formación. Es interesante señalar que los datos de micromatrices indican que, además de en polen, el gen 2 se expresa en etapas intermedias del desarrollo de la semilla, momento en el que se da la formación de la cubierta (Haughn y Chaudhury, 2005). No se han observado diferencias entre las cubiertas de las semillas mutantes para el gen 2 y las silvestres, pero hay que tener en cuenta que la cubierta procede de tejidos maternos, por lo que será heterocigota para la inserción. La cubierta protectora de la semilla presenta características comunes con la superficie del grano de polen. Resulta tentador sugerir que la proteína ATANKTM2 pudiera estar implicada en la formación de ambas estructuras, aunque para ello primero habría que determinar en que tejidos de la semilla se expresa el gen.

## II.5 Posibles funciones de los genes *AtAnkTm*

Al inicio de este trabajo, sólo uno de los genes *AtAnkTm* había sido estudiado (*ACD6*, 11). Este gen codifica una proteína involucrada en la ruta del ácido salicílico en respuesta de defensa (Lu *et al.*, 2003), pero la función de la proteína a nivel molecular permanece desconocida. Se han encontrado proteínas con similar organización de dominios en otras especies vegetales (dicotiledóneas y monocotiledóneas) pero sus funciones también permanecen desconocidas. También se encontró una organización similar de dominios en

algunas proteínas animales cuyas funciones son conocidas y pueden sugerir algunos papeles para los genes *AtAnkTm* (Figura 58).

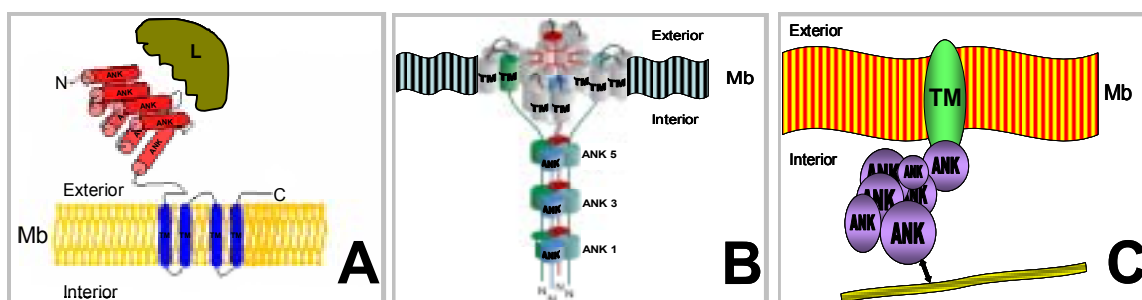


Figura 58.- Esquema representativo de las posibles funciones de los genes *AtAnkTm*. (A) Receptor de membrana (ej. proteína ANKTM1 humana), (B) canales de membrana (ej. proteína TRPV humana) y (C) anclaje a membrana (ej. unión entre anquirina/espectrina y citoesqueleto). Abreviaturas: Ank, repetición anquirina; TM, dominio transmembrana; Mb, membrana; N y C, extremos terminales de la proteína; L, ligando.

#### (a) Receptores de membrana.

La localización en membrana y la capacidad de las repeticiones ANK de unirse a otras proteínas sugieren que una de las posibles funciones que realizan sea la de receptores (Figura 58.A). Para ello la proteína ANKTM debería estar localizada preferentemente en la membrana citoplasmática con las repeticiones ANK encaradas hacia el exterior de la célula. Existen numerosos ejemplos de proteínas similares en animales que cumplen esta función:

- La proteína ANKTM1 es un canal activado por frío y mentol (Story *et al.*, 2003)
- Los receptores Notch (MacKenzie *et al.*, 2004).
- Receptor de neurotrofinas ARMS+kidins220 (Arevalo *et al.*, 2004).

#### (b) Canales de membrana.

La localización en membrana y la capacidad de asociarse entre sí permiten imaginar que las proteínas ANKTM puedan realizar funciones de facilitar el paso de sustancias a través de una membrana. De hecho, existen ejemplos en animales (Figura 58.B):

- Canales humanos de entrada de calcio CaT1 y CaT2 (Peng *et al.*, 2001).
- Canal de cationes OTRPC4 (Strotmann *et al.*, 2000).

- Discusión -

- Canal potencial de receptor transiente presente en las células receptoras de sabor.
- Familia de canales TRPV (Arniges *et al.*, 2006).

Estos canales estarían formados por varios polipéptidos ANKTM iguales o no, por un lado anclados en la membrana por los dominios transmembrana, y por otro, unidos entre sí por las repeticiones ANK (revisión en Niemeyer, 2005). En este caso, la localización de las proteínas puede ser en cualquier membrana celular que implique tráfico de sustancias y la orientación de las repeticiones ANK puede ser cualquiera. La función de estos complejos como canales no es incompatible con la función anterior de receptores.

Algunas de las proteínas AtANKTM tienen cierta similitud de secuencia en la región de los dominios transmembrana con la proteína TM20 de maíz (Figura 59). TM20 es una proteína necesaria para el desarrollo normal del embrión y contiene veinte segmentos transmembrana que pueden ser agrupados en 5 repeticiones formadas por cuatro segmentos (Stiefel *et al.*, 1999). Datos recientes indican que TM20 actuaría como transportador de auxinas (Jahrman *et al.*, 2005). En *Arabidopsis* ningún gen codifica para una proteína con 20 dominios transmembrana. Una posibilidad podría ser que en *Arabidopsis*, la función de la proteína TM20 de maíz sea llevada a cabo por un complejo de proteínas ATANKTM unidas por las repeticiones ANK.

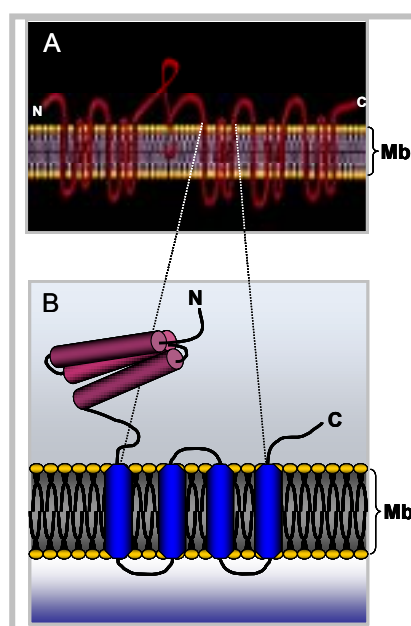


Figura 59.- (A) Modelo hipotético de la proteína transmembrana TM20 de maíz. La predicción de TM20 indica la presencia de 5 grupos con cuatro dominios transmembrana cada uno. (B) Modelo hipotético de la proteína ATANKTM codificada por el gen 15 (At4g11000) de *Arabidopsis thaliana*. Cilindros azules representan los dominios transmembrana. Cilindros granates corresponden a los dominios anquirina. Líneas punteadas entre (A) y (B) representan la región de la proteína que presenta mayor similitud entre TM20 y ATANKTM15. Abreviaciones: Mb, membrana; N y C, extremos terminales de las proteínas. N- y C-terminales de las proteínas TM20 y ATANKTM15 están orientados hacia el interior de la célula.

(c) Proteínas de anclaje a membrana.

Podrían servir para unir otras proteínas a la membrana citoplasmática. Esta función sería similar a la que realiza la proteína Anquirina al anclar el citoesqueleto a la membrana (Bennett, 1992; Bennett y Baines, 2001). La proteína anquirina es una proteína animal que se caracteriza por contener 24 repeticiones ANK (el nombre de las repeticiones proviene de esta proteína). La proteína anquirina se une, por una parte, a la proteína espectrina, un componente del citoesqueleto, y por otro, a proteínas de membrana (Figura 58.C) (Bennett y Baines, 2001). De esta manera se consigue anclar el citoesqueleto a la membrana. Las proteínas ANKTM podrían realizar una función parecida pero sin necesidad de unirse a otras proteínas de membrana ya que ellas mismas ya estarían localizadas en este orgánulo celular. En este caso la localización de la proteína podría ser cualquier membrana celular pero en especial la membrana citoplasmática, y la orientación de las repeticiones ANK debería ser hacia el citoplasma, aunque podrían existir otras alternativas como por ejemplo el interior del núcleo.

En resumen, la posición de las repeticiones anquirina hacia fuera o dentro de la membrana y la definición de la membrana en la que se localiza la proteína (citoplasmática, nuclear o de algún otro orgánulo celular) permitirían sugerir algunas de las posibles funciones de las proteínas ATANKTM en vegetales. Mediante la utilización de programas informáticos es posible estimar las características de determinadas proteínas en cuanto a la posición de los fragmentos transmembrana y de la ubicación que tendría el resto de la proteína, ya sea hacia el interior o exterior de la célula u orgánulo celular. Uno de estos programas es TMHMM v. 2.0 (<http://www.cbs.dtu.dk/services/TMHMM/>; Center for Biological Sequence Analysis, Technical University of Denmark DTU) que permite predecir las hélices transmembrana en proteínas. Al analizar la posición de la región anquirina de todas las proteínas ATANKTM se encontró que un 60% de ellas poseen dicha región orientada hacia el interior de la célula u orgánulo. Todas las repeticiones ANK codificadas por los genes *AtAnkTm* de la familia IV están orientadas hacia el interior de la célula u orgánulo y gran parte la familia I y toda la familia VI poseen la región anquirina orientada hacia el exterior de la

célula u orgánulo. Esto indicaría además, junto con los resultados de patrón de expresión de los genes *AtAnkTm*, que las proteínas codificadas por estos genes estarían participando en diferentes tipos de procesos a nivel celular.

## II.6 Estudio del gen *AtAnkTm28* durante la embriogénesis de *Arabidopsis thaliana*.

Uno de los objetivos de esta tesis es el de identificar genes de expresión específica en etapas iniciales del desarrollo de la semilla de *Arabidopsis*. Uno de los genes *AtAnkTm* (28) estaría dentro de esta categoría. El gen 28 se expresa durante los estadios iniciales del desarrollo de la semilla. Mediante hibridación *in situ* hemos visto que dicha expresión es especialmente abundante en el embrión durante los estadios iniciales (globular) de la embriogénesis, aunque también se expresa en el suspensor y el endospermo. La proteína parece estar localizada únicamente en ciertos lugares muy concretos de la membrana citoplasmática. El análisis mediante el programa TMHMM v.2.0, indica que la región N terminal de la proteína (que contiene las repeticiones ANK) estaría orientada hacia el interior de la célula (Figura 59).

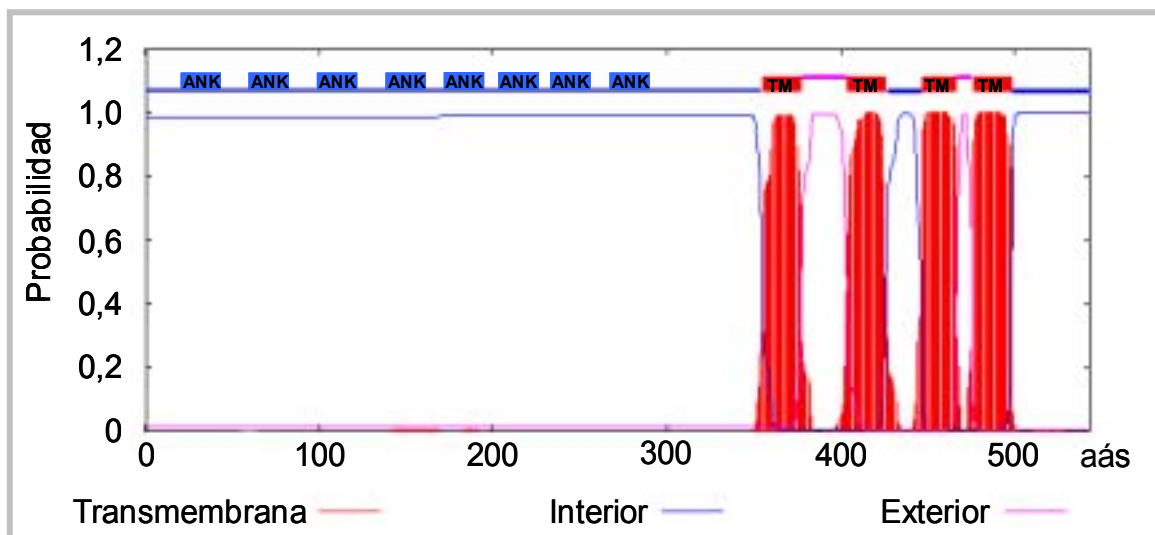


Figura 59.- Representación esquemática de la distribución de dominios transmembrana (rojo) en la proteína ATANKTM codificada por el gen EN28 de *Arabidopsis thaliana*. Rectángulos azules representan dominios anquirina; rectángulos rojos representan los dominios transmembrana; aás, aminoácidos. Predicción obtenida a partir del programa TMHMM v. 2.0 (<http://www.cbs.dtu.dk/services/TMHMM/>)



Los datos obtenidos no nos permiten determinar la función concreta de este gen. Cualquiera de las tres posibles funciones anteriormente citadas podría ser aplicable al gen 28. Los datos de localización subcelular indican que la región C-terminal de la proteína, que contiene los dominios transmembrana, se localiza en puntos concretos de la membrana. Esta localización parecería más bien indicar que la función de esta proteína tiene que ver con el anclaje a membrana más que a receptores o canales, aunque ninguna de las posibilidades puede ser descartada. Los puntos donde se localiza la proteína podrían corresponder a diferentes estructuras celulares como por ejemplo, plasmodesmos.

## **CONCLUSIONES**

---

1. La obtención de ESTs de semilla inmadura de *Arabidopsis* ha permitido determinar las categorías funcionales predominantes de los genes que se expresan en semillas entre los estadios globular medio y cotiledón curvado – verde temprano.
2. Se ha demostrado experimentalmente que el análisis *in silico* combinado de los bancos de ESTs y las bases de datos de hibridaciones de micromatrices permite identificar genes en *Arabidopsis* que se transcriban de manera específica o al menos predominante en un determinado órgano o estadio de desarrollo.
3. La representación de las diferentes categorías funcionales de los genes de expresión específica durante las etapas iniciales del desarrollo de la semilla difiere significativamente respecto a los presentes en el genoma completo y también respecto a los genes presentes en genotecas no seleccionadas de cDNA de semillas inmaduras.
4. El genoma de *Arabidopsis* contiene 105 genes que codifican proteínas con repeticiones anquirina que contienen un total de 509 de estas repeticiones. Estos genes se pueden separar en 16 grupos de acuerdo a los dominios identificados en las proteínas.
5. La secuencia consenso para la repetición anquirina en *Arabidopsis* posee gran similitud con la secuencia consenso descrita en animales.
6. El genoma de *Arabidopsis* contiene 37 genes (el grupo más abundante de los 16 identificados), que codifican proteínas con repeticiones anquirina y dominios transmembrana (genes *AtAnkTm*). Los análisis filogenéticos demuestran que los genes *AtAnkTm* se reparten en 6 familias. Otros tres genes que codifican proteínas únicamente con repeticiones anquirina están filogenéticamente relacionados con ellos.
7. Los genes *AtAnkTm* presentan patrones de expresión muy diferentes, incluso dentro de una misma familia, lo cual sugiere que cumplen funciones muy variadas.

8. La mutación de los genes *AtAnkTm2* y *AtAnkTm9* produce una reducción en la viabilidad de los granos de polen. La mutación de *AtAnkTm2* también produce cambios en la estructura de la exina.
9. El gen *AtAnkTm28* se expresa específicamente en embrión en estadio temprano de desarrollo. La fusión de la región transmembrana de ANKTM28 a GFP se localiza en algunos puntos de la membrana plasmática.
10. La búsqueda en *Arabidopsis* de proteínas semejantes a TM20 de maíz fue infructuosa. Es posible que la interacción de proteínas ATANKTM mediante las repeticiones anquirina permita formar complejos que cumplan funciones que en otras especies son realizadas por proteínas de mayor tamaño como TM20.

## **MATERIALES Y MÉTODOS**

---

## 1. MATERIALES

### 1.1 MATERIAL VEGETAL

#### 1.1.1. Especies empleadas y condiciones de crecimiento

En este trabajo se han utilizado plantas de *Arabidopsis thaliana* del ecotipo Columbia 0 (Col-0) y plantas del cultivar Gaspard de Colza (*Brassica napus*). Las líneas mutantes de inserción de T-DNA pertenecen al ecotipo Col-0.

Las semillas de *Arabidopsis* y colza no requieren esterilización previa cuando son sembradas sobre sustrato. Las semillas se sembraron directamente en macetas de polietileno con sustrato y fueron sometidas durante 48 horas a un período de vernalización para romper la latencia de las semillas y conseguir una germinación homogénea. Las plantas de *Arabidopsis thaliana* y colza germinaron y crecieron en el invernadero bajo las siguientes condiciones: 18 – 22 °C de temperatura, 45% de humedad, 16 h luz/ 8 h oscuridad.

#### 1.1.2. Mutantes

Las líneas mutantes de inserción de T-DNA utilizadas (Cuadro 18) se sembraron y crecieron utilizando las mismas condiciones bajo invernadero que el ecotipo Columbia – 0 de *Arabidopsis thaliana*. Todas las líneas fueron evaluadas a nivel de fenotipo (características visuales) con énfasis en el desarrollo de las semillas.

Las líneas mutantes SAIL asociadas a algunos de estos genes, poseen además de la inserción T, un gen de resistencia al herbicida BASTA. Por lo tanto, se sembró y creció una primera generación de cada una de estas líneas y seleccionaron con la aplicación del herbicida. Las semillas de las líneas SAIL fueron aportadas por Syngenta (Torrey Mesa Research Institute, EEUU). También se analizaron líneas mutantes de inserción de T-DNA de los bancos de semillas SM (Exon trapping Insert Consortium), Salk (The Salk Institute for Biological Studies) y WiscDsLox (University of Wisconsin);

Cuadro 18. Líneas mutantes de inserción de los genes *AtAnkTm*.

Código Atg	Líneas mutantes
At1g03670	Salk_070803; Salk_095446
At4g03440	SAIL_63
At4g03450	SM_3_31790
At4g03460	SM_3_21515; SM_3_2835
At4g03480	Salk_007058; Salk_063147
At4g03500	WiscDsLox442C8
At4g05040	SAIL_642
At4g11000	SAIL_44
At5g15500	Salk_019783; Salk_071042; SAIL_66
At5g51160	Salk_084573
At5g54620	Salk_053630
At5g50140	SM_3_32888; SM_3_32883
At1g05640	Salk_007101
At1g07710	Salk_008522; Salk_008523; Salk_043439
At3g09550	SAIL_633
At5g02620	SAIL_20
At5g60070	SAIL_140; SM_3_15526; SM_3_15534
At5g04690	SM_3_16667

### **1.1.3. Esterilización de semillas de Arabidopsis**

Las semillas (1.000–1.500) se depositan en un tubo de microcentrífuga abierto y se coloca en un desecador que contiene un recipiente con 100 ml de lejía. Se añaden 3 ml de HCl 37% (es conveniente realizar este tratamiento en una campana de gases). Se sella inmediatamente el desecador y se mantiene durante un mínimo de 4 horas (No es conveniente excederse en el tiempo pues las semillas podrían no germinar).

### **1.1.4. Cultivo *in vitro***

#### Preparación de Medio

1. Agregar 4,4 g de sales y vitaminas Murashige & Skoog (MS; Duchefa) a 1 botella que contiene 1 l de agua (Milli Q).
2. Colocar en agitación con barra magnética y añadir de sacarosa 10 g/l y de MES 0,5 g/l.
3. Ajustar el pH a 5,7 – 5,8 con KOH.
4. Adicionar 8 g/l de agar (Difco) y poner en agitación durante algunos minutos. Retirar la barra magnética.
5. Autoclavar
6. Atemperar a 50 – 55 °C y plaquear bajo cabina de flujo laminar.

#### Obtención de raíces de *Arabidopsis thaliana*

1. Se necesitan plantas de 2 hojas verdaderas (o después de una semana de sembrar en placas Petri con medio Gambourg B5 *Medium* (GM)).
2. Colocar 20 - 25 plántulas en un matraz Erlenmeyer de 500 ml con 200 ml de medio B5 líquido a una temperatura de 22 °C, en agitación de 130 - 150 rpm, durante un período de dos semanas.

#### Preparación de medio líquido B 5 (Gambourg)

- Agregar 3,16 g de sales y vitaminas B 5 (Duchefa) a 1 botella que contiene 1 l de agua (Milli Q).
- Colocar en agitación con barra magnética y añadir 20 g/l de glucosa y 0,5 g/l de MES.
- Ajustar el pH a 5,7 – 5,8 con KOH.
- Autoclavar

## **1.2. CEPAS BACTERIANAS Y VECTORES DE CLONAJE**

Para el clonaje, amplificación y análisis de DNA se utilizaron los siguientes vectores y cepas bacterianas:

#### Vectores utilizados para la clonación de productos de PCR:

pTZ57R/T (Fermentas); pGEM<sup>®</sup>-T Easy (Promega); pCR<sup>®</sup>II (Invitrogen).

#### Vectores utilizados para construcciones génicas:

pUC1303: Plásmido derivado del pUC18 y el pCAMBIA1303.

#### Cepas bacterianas:

*Escherichia coli* K-12 DH5 $\alpha$ ;

*One shot<sup>®</sup> TOP10 competent cells* (Invitrogen).

## 2. MÉTODOS

---

### 2.1. EXTRACCIÓN DE ÁCIDOS NUCLEICOS

#### 2.1.1. Extracción de DNA plasmídico

##### 2.1.1.1 Minipreparaciones de DNA plásmidico

El rendimiento que se obtiene en las extracciones de DNA plasmídico dependen tanto del tamaño y tipo plásmido como del tamaño del inserto. Es importante conocer si se trabaja con un plásmido de alto o bajo número de copias para adaptar la metodología.

1. Preparar un inóculo con 3 ml LB y 3µl carbenicilina e incubar a 37°C durante toda la noche (o/n) en agitación (250 rpm).
2. Transferir 1,5 ml del cultivo a un tubo de microcentrífuga, centrifugar 4 min a temperatura ambiente (Ta) y descartar el sobrenadante.
3. Transferir el resto del cultivo y repetir el paso anterior.
4. Resuspender el pellet en 190 µl de SOLUCIÓN I fría utilizando el vórtex.
5. Incubar 5 min en hielo.
6. Añadir 400 µl de SOLUCIÓN II (preparada en el mismo momento), invertir 5 veces (no vortear) e incubar 5 min en hielo.
7. Añadir 300 µl de SOLUCIÓN III fría, invertir 5 veces (no agitar) e incubar en hielo mínimo 10 min.
8. Centrifugar 4 min a Ta y transferir 750 µl de sobrenadante a un nuevo tubo. Añadir 1 V de fenol/cloroformo/alcohol isoamílico (25:24:1), vortear y centrifugar 1 min; transferir 725 µl de la fase superior a un nuevo tubo.
9. Añadir 1 V cloroformo/alcohol isoamílico (24:1), agitar y centrifugar 1 min; transferir la fase superior a un nuevo tubo.
10. Añadir 0.6 V de isopropanol (435 µl) a Ta, invertir e incubar 10 min a Ta.
11. Centrifugar 10 min a Ta y descartar el sobrenadante.
12. Lavar el pellet con 200 µl de EtOH 70%.
13. Centrifugar 5 min a Ta y eliminar el sobrenadante.
14. Secar el pellet y resuspender en el volumen deseado de TE pH 8.0. Para facilitar la resuspensión incubar 10 min a 65°C.
15. Añadir 1 µl RNasa e incubar 15 min a 37°C.

Solución I: glucosa 50 mM, Tris-HCl 25 mM pH 8.0, EDTA 10 mM

Solución II: NaOH 0.2 N, SDS 1%

Solución III: KAc 3 M pH 5.5

Medio LB: Bactotripton 10g/l, extracto de levadura 5g/l, NaCl 5g/l. Ajustar a pH 7.0 con NaOH y autoclavar.

##### 2.1.1.2. Preparaciones a gran escala

En los casos que ha sido requerida una gran cantidad de DNA se ha utilizado el kit comercial de QIAGEN, que se basa en la lisis alcalina y el uso de columnas con una matriz de intercambio iónico. Se consigue así, un DNA de gran pureza y libre de actividad DNasa que podrá ser empleado en cualquier



tipo de reacción. Normalmente se han utilizado midipreparaciones con 50 ml de cultivo.

### **2.1.2. Extracción de DNA genómico**

1. Homegenizar entre 100 y 200 mg de tejido en mortero con N<sub>2</sub> líquido y transferir a un tubo de microcentrífuga.
2. Añadir 600 µl de tampón de extracción y 50 µl de SDS 20%.
3. Vortear a alta potencia durante 2 min.
4. Incubar a 65 °C durante 10 min.
5. Añadir 2 µl de RNAsa e incubar 10 min a 37 °C.
6. Añadir 230 µl de acetato potásico 5M y agitar.
7. Centrifugar 10 min a 4°C y transferir el sobrenadante a un nuevo tubo.
8. Añadir 1 V de fenol/cloroformo/alcohol isoamílico (25:24:1).
9. Mezclar por inversión y centrifugar 5 min a 13.000 rpm.
10. Recuperar la fase acuosa y volver a repetir la extracción.
11. Añadir 360 µl de isopropanol frío (-20°C) y centrifugar 5 min 13.000 rpm.
12. Descartar el sobrenadante y lavar el pellet con 200 µl de EtOH 70% frío (-20°C).
13. Dejar secar y resuspender en 50 µl de TE (pH 8).

Tampón de extracción: Tris-HCl 100mM (pH 8.0), EDTA 50mM (pH 8.0), NaCl 500mM y β-Mercaptoetanol 10mM.

### **2.1.3. Extracción de RNA**

Es importante tener presente siempre la fragilidad del RNA y evitar la existencia de RNAsas contaminantes.

#### **2.1.3.1. Método del LiCl**

Método descrito por Vicient y Delseny (1999) especialmente indicado para la extracción de RNA total a partir de semillas de Arabidopsis y otras plantas relacionadas como brasicáceas, tabaco, encina, almendro o patata. Se puede utilizar también para silicua inmadura, hoja, tallo, flor o polen.

1. Homogenizar 100 mg de tejido en un mortero con N<sub>2</sub> líquido.
2. Transferir el material a un tubo de microcentrífuga que contenga 1.8 ml de tampón de extracción previamente enfriado a -20°C.
3. Resuspender mediante agitación en vórtex. Incubar O/N a 4°C.
4. Centrifugar 13.000 rpm durante 4 seg y transferir el sobrenadante a un nuevo tubo.
5. Centrifugar a 13.00 rpm durante 30 min a 4°C y descartar el sobrenadante.
6. Lavar el pellet con 500 µl de EtOH 70% frío (4°C) y centrifugar 1 min a 4°C.
7. Secar el pellet y resuspender en 1 ml de tampón de disolución
8. Añadir un V de fenol equilibrado a pH 7,6. Agitar manualmente.
9. Centrifugar 15 min a 4°C. Pasar la fase superior a un nuevo tubo.
10. Repetir los pasos 13 y 14.
11. Repetir los pasos 13 y 14, esta vez con fenol/cloroformo/alcohol isoamílico (25:24:1).

12. Repetir los pasos 13 y 14, con cloroformo/alcohol isoamílico (24:1).
13. Repartir la última fase acuosa en 2 tubos y añadir 0.1 V NaAc 3 M y 1.5 V EtOH 100%. Centrifugar 20 min a 4°C.
14. Eliminar el sobrenadante y lavar el pellet con 500 µl de EtOH 70% frío (4°C). Centrifugar 5 min a 4°C.
15. Eliminar el etanol y dejar secar a temperatura ambiente.
16. Resuspender en 50 ó 100 µl de H<sub>2</sub>O libre de RNAsas.

Tampón de extracción: LiCl 8M, β-Mercaptoetanol 2%.

Tampón de disolución: SDS 0.5%, NaCl 100mM, EDTA 25 mM, TrisHCl 10 mM (pH 7.5), β-Mercaptoetanol 2%.

### 2.1.3.2. Método del Trizol

Este método está basado en el uso del reactivo Trizol<sup>R</sup> (Invitrogen) que minimiza el tiempo invertido en el proceso de extracción. El protocolo seguido ha sido el descrito por el fabricante.

1. Homogenizar 50- 100 mg de tejido en un mortero con N<sub>2</sub> líquido.
2. Transferir el material a un tubo de microcentrifuga. Añadir 1 ml de Trizol<sup>R</sup> y mezclar por inversión.
3. Incubar 5 min a temperatura ambiente.
4. Añadir 0.2 V cloroformo, mezclar por inversión e incubar 2-3 min a Ta.
5. Centrifugar a 12.000 x g durante 15 min a 4°C. Transferir la fase acuosa a un nuevo tubo.
6. Precipitar el RNA con 0.5 ml de isopropanol e incubar 10 min a Ta.
7. Centrifugar 12.000 x g 10 min a 4°C y eliminar el sobrenadante.
8. Lavar el pellet con 1 ml EtOH 70%, agitar y centrifugar a 7.500 x g 5 min a 4°C.
9. Eliminar el sobrenadante y dejar secar el pellet.
10. Resuspender en H<sub>2</sub>O libre de RNAsas e incubar 10 min a 60°C.

### 2.1.3.3. Tratamiento del RNA con DNAsa

El tratamiento del RNA total con DNAsa permite degradar el DNA genómico contaminante que no se ha podido eliminar completamente durante el proceso de extracción del RNA y que podría interferir en posteriores análisis y/o ensayos.

1. Añadir en un tubo de microcentrifuga los siguientes reactivos:

RNA total	50 µg
Tampón DNAsa 10X	10 µl
RNA guard <sup>TM</sup> (24.8U/µl)	2 µl
DNAsa RNAsa-free (10U/µl)	3 µl
H <sub>2</sub> O libre de RNAsas	Completar hasta 100 µl

2. Incubar 30 min a 37°C.
3. Añadir 100 µl de H<sub>2</sub>O libre de RNAsas y añadir 1 V de fenol/cloroformo/alcohol isoamílico (25:24:1) y mezclar por inversión.

4. Centrifugar 5 min a 13.000 rpm a 4°C y recuperar la fase acuosa.
5. Añadir 1 V de cloroformo/ alcohol isoamílico (24:1), mezclar por inversión y centrifugar 5 min a 13.000 rpm a 4°C.
6. Recuperar la fase acuosa y precipitar el RNA con 0.1 V NaAc 3M y 2.5 V EtOH 100%, incubar 1 hora a - 80°C.
7. Centrifugar 20 min a 13.000 rpm y eliminar el sobrenadante.
8. Lavar el pellet con 200 µl EtOH 70% y centrifugar 5 min a 13.000 rpm.
9. Secar el pellet y resuspender en el volumen necesario de H<sub>2</sub>O libre de RNasas para tener una concentración final de 1µg/µl.

Tampón DNAsa 10X: Tris-HCl 0.4 M (pH 7.5), MgCl<sub>2</sub> 60 mM.

*RNA guard*<sup>TM</sup> *RNAse Inhibitor Porcine* (24.8U/µl) (Amersham Bioscience)

*DNAsa I RNAse-free* (10 U/µl) (Roche)

## 2.2. ELECTROFORESIS DE ÁCIDOS NUCLEICOS

### 2.2.1. Electroforesis de DNA en gel de agarosa

Los fragmentos de DNA pueden separarse por electroforesis en geles de agarosa a un determinado porcentaje en función de su tamaño. El DNA está cargado negativamente de manera que migra siempre hacia el electrodo positivo.

Un gel de agarosa a 1% se prepara fundiendo 1g de agarosa en 100 ml de tampón TAE 1X. La solución se enfría hasta 55°C y se le añade bromuro de etidio, para la visualización de los ácidos nucleicos por luz UV en un transiluminador. Antes de cargar el gel se añade tampón de carga a las muestras y la electroforesis se lleva a cabo en tampón TAE 1X y un voltaje 50-90 Voltios durante el tiempo que sea preciso.

TAE 1X: Tris Base 40 mM, HAc 20 mM, EDTA 2 mM pH8.

Tampón de carga 10X: azul de bromofenol 0.25%, xylene cyanol FF 0.25%, glicerol 50%.

### 2.2.2. Electroforesis de RNA en geles de formaldehído/agarosa

Debido a la facilidad de degradación del RNA es conveniente realizar la electroforesis en gel de agarosa en condiciones desnaturalizantes. De forma general se ha utilizado el formaldehído, sustancia desnaturalizante tóxica y cancerígena por inhalación y contacto que debe utilizarse con cuidado.

Se funde la agarosa en tampón MEN 1X y se atempera antes de añadir el formaldehído a una concentración final de 6.5%. Se añade tampón de carga a las muestras de RNA y se desnaturalizan 5 min a 65°C antes de cargar el gel. La electroforesis se realiza en tampón MEN 1X y a 40 - 90 Voltios durante el tiempo que sea necesario.

MEN 10X: MOPS 0.2 M (pH 7), NaAc 50 mM, EDTA 10 mM (pH 8). Ajustar a pH 7 y autoclavar. Almacenar a 4°C.

Tampón de carga 10X: MEN 1X, formaldehído 6.5%, formamida 50%, azul de bromofenol 0.6X.

### **2.2.3. Recuperación de DNA a partir de gel de agarosa**

Existen diversos métodos de recuperación de fragmentos de DNA de geles de agarosa. Se han utilizado distintos y variados kits comerciales que se basan en el uso de columnas con una matriz de silica-gel, y de un pH y una concentración de sales favorables. Durante este trabajo se ha utilizado, entre otros, el kit comercial QIAquick Gel Extraction de la casa Qiagen siguiendo el protocolo descrito por el fabricante.

1. Colocar en un tubo de microcentrífuga el fragmento recortado del gel donde se localiza el DNA.
2. Añadir 300  $\mu$ l de tampón QG por cada 100  $\mu$ g de gel.
3. Incubar 10 min a 50°C, mezclando con vórtex cada 2-3 min.
4. Añadir 100  $\mu$ l de isopropanol por cada 100  $\mu$ g de gel, solamente si el fragmento de DNA es menor de 500 pb o bien mayor de 4 kb.
5. Colocar la columna en un tubo de 2 ml, añadir la muestra y centrifugar 1 min a 13.000 rpm. Descartar la elución.
6. Lavar la columna con 750  $\mu$ l de tampón PE y centrifugar 1 min a 13.000 rpm. Descartar la elución y colocar la columna en un tubo nuevo.
7. Eluir el DNA con 50  $\mu$ l de tampón EB. Incubar 1 min y centrifugar 1 min a 13.000 rpm.

Tampón QG : suministrado por el kit, contiene tiocianato de guanidinio.

Tampón PE: suministrado por el kit.

Tampón EB: Tris-HCl 10 mM (pH 8.5)

## **2.3. MODIFICACIONES GENERALES DEL DNA**

### **2.3.1. Digestión con enzimas de restricción**

La digestión con enzimas de restricción de un plásmido nos permitirá tanto la liberación del inserto o fragmento de DNA clonado en dicho plásmido así como su posterior clonación dirigida en otros vectores ya sean de expresión o bien, de transcripción *in vitro*, etc.

Una unidad de enzima es la cantidad necesaria de enzima para digerir 1  $\mu$ g de DNA durante 1 hora a 37°C, en el tampón apropiado para cada enzima.

De forma general, se ha digerido con 2-3 U de enzima por  $\mu$ g de DNA en un volumen final de reacción de 20  $\mu$ l. La concentración de glicerol que contiene el enzima no debe sobrepasar el 10% del volumen final de la reacción. Los enzimas que han sido empleados durante este trabajo pertenecían a las casa comerciales Promega o Roche.

## 2.4. SUBCLONAJE DE FRAGMENTOS DE DNA EN PLÁSMIDOS

### 2.4.1. Ligación de fragmentos de DNA a un vector de clonación

La reacción de ligación por regla general se realizado con 200 ng de DNA en un volumen final de 10  $\mu$ l. La cantidad de inserto y de vector que se requieren para la reacción se calcula según la ecuación (Sambrook, 1989):

Ligación de extremos cohesivos:

$$[\text{ng inserto/ kb inserto}] = 3 \times [\text{ng plásmido/ kb plásmido}]$$

Ligación de extremos romos:

$$[\text{ng inserto/ kb inserto}] = 5 \times [\text{ng plásmido/ kb plásmido}]$$

A la cantidad de inserto y vector estimados se añade el volumen necesario de tampón T4 DNA ligasa 10X y de 0.1-0.5 U de enzima T4 DNA ligasa por  $\mu$ l de reacción. La reacción se incuba 2 horas a  $T_a$ , o bien, o/n a 16°C.

En este trabajo se ha empleado mayoritariamente el enzima T4 DNA ligasa y el tampón T4 DNA ligasa 10X de la casa comercial Promega. Así como los vectores principalmente utilizados han sido el vector pGEM-Teasy de Promega, el vector pTZ57R de Fermentas y vector pCRII-TOPO de Invitrogen.

Para la clonación de productos de PCR en el vector pCRII-TOPO el proceso de ligación no requiere la presencia de DNA ligasa y puede llevarse a cabo en 5 min con una elevada eficiencia. Esto es gracias a la actividad de la DNA topoisomerasa I que tiene conjugada el vector en el extremos 3'OH.

### 2.4.2. Preparación de células competentes

La transformación de *E. coli* con vectores plasmídicos permite mantener de manera estable fragmentos de DNA exógeno y amplificarlos mediante el cultivo de las células transformadas. Para que el vector pueda ser introducido en una bacteria, ésta debe ser competente para la transformación. Existen una gran variedad de protocolos para preparar células competentes y para su transformación. A continuación se describen los métodos utilizados en esta memoria.

#### 2.4.2.1. Método del $\text{CaCl}_2$

1. Inocular 250 ml de medio de cultivo LB con 2.5 ml de precultivo de la cepa de interés.
2. Crecer a 37°C en agitación hasta que la densidad óptica llegue a  $\text{DO}_{550}=0.39-0.45$ .
3. Enfriar el cultivo en hielo durante 20 min. A partir de este momento trabajar siempre en hielo.
4. Transferir el cultivo a botellas de polipropileno de 250 ml y centrifugar 1.000 x g durante 10 min a 4°C.
5. Descartar el sobrenadante, secar bien el interior de la botella y resuspender las células en 50 ml de  $\text{CaCl}_2$  50 mM frío.

6. Dejar en hielo de 10 a 60 min.
7. Transferir a tubos COREX de 30 ml y centrifugar 1.000 x g durante 10 min a 4°C.
8. Decantar el sobrenadante, secar y resuspender las células en 10 ml de CaCl<sub>2</sub> 0.1 M frío.
9. Mantener en hielo de 2 a 24 horas.
10. Para almacenar las células es necesario previamente añadir 0.2 V de glicerol frío y posteriormente alicuotar en 100 µl.
11. Almacenarlas a – 80°C.

#### **2.4.2.2. Para electroporación**

1. Inocular 2 ml de LB con la cepa de interés e incubar a 37°C o/n, en agitación a 250 rpm.
2. Inocular 200 µl del cultivo en 100 ml (ó 500 ml) de LB e incubar a 37°C, a 200 rpm hasta que la DO<sub>600nm</sub> sea de 0.5–0.6.
3. Transferir a una botella 250 ml de polipropileno estéril. Mantener el cultivo en hielo.
4. Centrifugar a 5.000 rpm, 10 min a 4 °C. Decantar sobrenadante.
5. Resuspender en 100 ml de H<sub>2</sub>O estéril fría y centrifugar a 5.000 rpm, 10 min a 4°C.
6. Resuspender en 50 ml H<sub>2</sub>O estéril fría y centrifugar a 5.000 rpm, 10 min 4°C.
7. Resuspender en 2 ml H<sub>2</sub>O estéril fría y repetir la centrifugación.
8. Resuspender en 2 ml de glicerol 10% frío y centrifugar a 5.000 rpm, 10 min a 4°C.
9. Resuspender en 0.3 ml de glicerol 10% frío.
10. Alicuotar 40 µl en tubos de microcentrífuga, mantener en N<sub>2</sub> líquido y almacenar a – 80°C.

#### **2.4.3. Transformación de células competentes de *E. coli***

##### **2.4.3.1. Transformación por choque térmico**

1. Descongelar la alícuota de células competentes para choque térmico en hielo.
2. Añadir 10-20 ng de plásmido y mezclar con pipeta.
3. Incubar 30 min en hielo.
4. Inducir el choque térmico incubando 90 s a 42°C. Transferir a hielo inmediatamente y mantener 2 min.
5. Añadir 800 µl de medio SOB o LB, e incubar 60 min a 37°C, en agitación.
6. Centrifugar 2 min a 6.000 rpm y resuspender el pellet en 200 µl del sobrenadante.
7. Plaquear en placas de LB-agar que contengan el antibiótico adecuado para la selección así como IPTG-X-Gal (4µg/ml). Para la selección de recombinantes en pCRII-TOPO se puede usar carbenicilina (100 µg/ml) y kanamicina (25 µg/ml) indistintamente.
8. Incubar las placas en posición invertida a 37°C o/n.
9. Seleccionar las colonias blancas para el análisis de la presencia de DNA de interés.

Medio SOB: bacto triptona 20g/l, extracto de levadura 5g/l, NaCl 0.5g/l. Ajustar a pH 7.0 y autoclavar. Añadir 10 ml de MgCl<sub>2</sub> estéril.

LB-agar: bacto triptona 10g/l, extracto de levadura 5g/l, NaCl 0.5g/l, agar 15g/l. Ajustar a pH 7 con NaOH.

### 2.4.3.2. Transformación por electroporación

La transformación con alto voltaje es uno de los métodos más eficientes de transformación de *E. coli*. Antes de proceder a la transformación es necesario eliminar la presencia de sales en la ligación pues podría provocar una sobrecarga en el momento de aplicar la descarga eléctrica durante la electroporación. El protocolo descrito a continuación es una modificación del propuesto por Sheen (1997).

#### Pretratamiento de la ligación

1. Incubar la ligación 10 min a 65°C.
2. Añadir 0.1 V NaAc 3M (pH 5.5) y 2.5 V EtOH 100%, vortear y centrifugar 20 min a 13.000 rpm.
3. Eliminar el sobrenadante y lavar el pellet con 200 µl EtOH 70%.
4. Centifugar 5 min a 13.000 rpm, eliminar el sobrenadante y secar el pellet.
5. Resuspender en el volumen deseado de H<sub>2</sub>O estéril.

#### Electroporación

1. Descongelar las células competentes para electroporación en hielo y añadir 10-20 ng de ligación. La eficiencia de transformación disminuye al excederse en la cantidad de plásmido.
2. Mantener 5 min en hielo.
3. Transferir las células a la cubeta y repartirlas uniformemente evitando la formación de burbujas.
4. Proceder a la descarga entre 1.8–2.5 kVoltios en el electroporador (BioRad).
5. Inmediatamente añadir 460 µl de SOB, mezclar y transferir las células a un tubo de microcentrífuga.
6. Incubar durante 1 hora a 37°C, en agitación.
7. Concentrar las células y plaquear en placas LB-agar como en el apartado 2.4.3.1

## 2.5. SECUENCIACIÓN

La secuenciación de DNA se ha realizado según el método descrito por Sanger (1977) en el Servicio de Secuenciación Automática del IBMB-CSIC de Barcelona. Se utilizan oligonucleótidos marcados con fluorocromos (ALF, *Automated Laser Fluorescent*, Pharmacia). El DNA secuenciado procedía de minipreparaciones y se requiere unos 0.5 µg/µl. Es posible lograr la lectura eficiente de hasta 600 pb.

Las secuencias obtenidas han sido procesadas y comparada con las bases de datos NCBI (Nacional Centre for Biotechnology Information) o del EBI (European Bioinformatics Institute).

## **2.6. GENOTECA DE cDNA**

La construcción de esta genoteca de cDNA de semilla inmadura de *Arabidopsis thaliana* ha sido realizada empleando el kit comercial SMART™ cDNA Library Construction Kit de la casa Clontech siguiendo el procedimiento establecido por el fabricante pero incluyendo algunas modificaciones.

### **2.6.1. Síntesis de la primera cadena de cDNA**

El RNA total fue extraído según el protocolo descrito en el apartado 2.1.3.1. Seguidamente se procede a la síntesis del cDNA de cadena simple a partir del RNA total mediante retrotranscripción con una modalidad del enzima MMLV-RT que carece de actividad RNasa H pero mantiene la actividad polimerasa normal, de manera que puede sintetizar fragmentos de cDNA más largos que la enzima normal. Se consigue así tener cDNA con la secuencia no traducida de la región 5' (5'UTR) del mRNA.

### **2.6.2. Síntesis de cDNA de doble cadena**

La síntesis de cDNA completo de doble cadena puede realizarse por dos métodos: por *Long-Distance PCR (LD-PC)* o PCR de larga distancia con parámetros especiales para fragmentos de gran tamaño, o bien, por *Primer Extension*. El uso de uno u otro método dependerá de la cantidad de RNA inicial para la construcción de la genoteca (50 ng y 1 µg, respectivamente). Sin embargo, ambas metodologías generan un enriquecimiento en el número de cDNAs obtenidos.

### **2.6.3. Ligación del cDNA**

Aunque el kit ofrece los reactivos y componentes necesarios para la clonación de los cDNAs en vectores de tipo fago, en este trabajo se ha procedido a la clonación de los cDNAs en el vector pCRII-TOPO. De esta manera, se han omitido los pasos de digestión con enzima Sfil y de fragmentación del cDNA. La ligación de los cDNAs al vector y la posterior transformación y selección de colonias recombinantes se ha realizado según el proceso ya mencionado anteriormente en el apartado 2.4.3.1.

## **2.7.- REACCIÓN EN CADENA DE LA POLIMERASA (PCR)**

### **2.7.1. Observaciones generales**

La reacción en cadena de la polimerasa (PCR) permite la síntesis de fragmentos de DNA específicos a partir de un DNA molde, en presencia de dos oligonucleótidos de secuencia complementaria a las dos cadenas de la secuencia molde, respectivamente, y del enzima DNA polimerasa. La reacción se lleva a cabo en un termociclador bajo un programa de temperaturas controladas de desnaturalización inicial; desnaturalización, unión de los oligonucleótidos y extensión de la polimerasa dentro de un número de ciclos determinado, y finalmente la extensión final.



Esta metodología es aplicable a un gran número de estrategias experimentales. Las más frecuentes utilizadas durante la realización de este trabajo son la amplificación de genes a partir de DNA genómico, análisis de la expresión de genes por RT-PCR a partir de cDNA y la confirmación de la presencia de inserto en la clonación en plásmido.

Las condiciones habituales de la reacción serían:

DNA	10-100 ng
Tampón PCR	1X
MgCl <sub>2</sub>	1-5 mM
dNTPs	100 µM
Oligonucleótido 1	25 pmol
Oligonucleótido 2	25 pmol
Taq polimerasa	2.5U/50µl reacción
H <sub>2</sub> O	hasta 50 µl de volumen final

94°C	5 min	1 ciclo
94°C	30 s	30 ciclos
Y°C	30 s	
72°C	Z s	
72°C	10 min	1 ciclo

Y = depende de T<sub>m</sub> de los primers (T<sub>m</sub> ± 5)

Z = depende de longitud del fragmento; 1Kb/min

Tampón de PCR 10X: KCl 500 mM, Tris-HCl 200 mM (pH 8.6), Triton X-100 1%.

### **2.7.2. RT-PCR semicuantitativa**

La técnica de RT-PCR (*Reverse Transcriptase – Polymerase Chain Reaction*) está basada en la retrotranscripción del RNA, libre de contaminación de DNA, a cDNA mediante el enzima Transcriptasa Reversa y un oligonucleótido dT, el cual se une a la cola de poliA del mRNA. El cDNA de cadena simple obtenido a partir del RNA es utilizado como molde en la posterior amplificación por PCR. La RT-PCR ha sido realizada siguiendo procedimiento establecido en el kit Omniscript Reverse Transcriptase (Qiagen).

1. Mezclar:

RNA	50 ng – 2 µg
Oligo-dT (10 µM)	2 µl
H <sub>2</sub> O	Hasta un volumen de 14 µl

2. Desnaturalizar el RNA 5 min a 65°C y transferir inmediatamente a hielo.

3. Seguidamente añadir:

Tampón RT 10X	2µl
dNTPs 5mM c/u	2µl
RNAguard™ (26,8U/µl)	1µl
Omniscript-RT (4U/µl)	1µl

4. Mezclar e incubar durante 1 hora a 37°C.

5. Añadir 20 µl de H<sub>2</sub>O y guardar a -80°C.

Obtención de la primera cadena de cDNA:

PCR sobre el cDNA:

cDNA	100 – 250 ng
Tampón de PCR	1X
dNTPs	100 µM
MgCl <sub>2</sub>	1-5 mM
Oligonucleótido A	25 pmol
Oligonucleótido B	25 pmol
Taq polimerasa	2,5 U
H <sub>2</sub> O	Hasta un volumen final de 50 µl

Los parámetros cíclicos utilizados han sido:

94°C	5 min	1 ciclo
94°C	30 s	
Y°C	30 s	30 ciclos
72°C	Z s	
72°C	10 min	1 ciclo

Para detectar las posibles amplificaciones de DNAg contaminante se diseñaron oligonucleótidos que flanquearan uno o varios intrones. De esta manera Para obtener resultados semicuantitativos se limitó el número de ciclos para así mantener la amplificación dentro de la fase exponencial y evitar la saturación de la reacción. El gen de la actina se utilizó para equilibrar la cantidad de cDNA molde.

**2.7.3. Oligonucleótidos**

Los oligonucleótidos utilizados en esta memoria se resumen a continuación.

**Análisis de genes de expresión específica en semilla**

Para el estudio de los genes de expresión específica en semilla de *Arabidopsis thaliana* se diseñaron los oligonucleótidos que se indican a continuación.

Código Atg	Oligonucleótidos (5' a 3')
At5g09640	GACACACCAAACATCAGAACCG CTACTCATCATCCAAGGTCTCC
At5g22470	TATGCTCTCTTCCGGTTCTGG ATGGAACCAACCGTCCACAAGG
At5g45690	ACGATTGCGACTCCTCTAAACC GAACGGAGCCAATTTCTGCATC
At1g67100	GCTCATGAACCTCCTCAACACC CCCGATCCAAGTCTTTGGTTCC
At3g60730	TCAAGCTGTGGCGTTGAGAGTG GGTAAACGGAGAAGCCTCTTCC
At3g12203	GGCACTGATCTCTGATGAACAC TTCTGAACCATCCATGGTCTCC
At1g71691	GCTTGTCTTTCATCGGAATGGG TACGACAAGGCGTTTCAAAGGG
At2g43260	TTCCGGCTTGAACCATAACTGC TGAACCACCTTTTCTGCCTTCG
At1g68380	TGTTTTATGGCCGCCGATTCC TCCAAGTAAGCGTCTATTCCG
At4g14780	TCAAACCTCGCTTTGATCTCGC TTTACCACCTCCTTCATCTCC
Oligo (dT)	CCTAGGCCACTGTGGCCTTTTTTTTTTTTTT

### Análisis de genes *AtAnkTm*

A continuación, se entrega un detalle de los oligonucleótidos específicos diseñados para el análisis de expresión de cada uno de los genes.

Código Atg	EN	Secuencia de los oligonucleótidos (5'a 3')
At4g03440	2	CTTGGATTTGCTACGTCGTAGCC GAGATACTGCTCTCCACTCAGCC
At4g03450	3	CTTGTACACGGCGCTCAAAGGC ACCCTTTGGCGAACAAAGTGCAC
At4g03480	6	CAGATTCCGCTTCATGTGGCCGC CAGGCAGCTGTCTCCATTTGGCG
At4g05040	9	GCAGGTAACAATGACCTTGAAGGG CCTGCAGCAAACGTCATTGTGGC
At4g14400	11	GACAACTGACCGTGAAGTGAGG AGAGCCGCTACCACGAGAAGAGC
At1g14480	12	CGGGATGGATCCAGAGAATGAGCC GGAGGGCCATCTGATAAGTGGCTG
At1g14500	13	GCCAAAAAGATTCTGCTTCCACCG GCTGGAGTGCAGTTTGATAAGTGG
At4g10720	14	GCGAGAAAACCTTAACACATACGGG GGTTATGTAAGATATCTAGGGCGG
At4g11000	15	GCTAAACGTGTCAGGTTTCAGCCC GGAAGGTGAGCACTGAGAGATAGC
At5g15500	16	AACGCCGACGGACTTACAC ATTCCAAAAACCAAACCTACC
At5g51160	17	GGGGGTTGAAAAGAAGCTTTGCCG GCTTCGCTCGGAAACATAACCAGC
At5g54610	18	GGGTGGATGCAGAAAATGCGCG GGGCTGCAGTCTGAAAAGTGGC
At5g54620	19	GCGAGACTGCTCTACATATTGCGG GCGATCTCAAGAAGACAGATGCGG
At2g24600	22	CTTGAGCTTGTGAGGGAGAAGG CTCTTCCCGTAACGCGTACGG
At5g54710	25	GGCCAGAGTGCAAACATACGCC CACCACCAGGAGGGTTTATCCCG
At1g07710	28	GGGGAACAGAACCAGTCAGGCG CCATGTGAAGGGCTGTTGGCCC
At2g01680	29	CGTTTTTCATGTCGCTGCCAAGCG CCGTAAGGTGACGCAATAAGCTCG
At3g09550	31	GTTGCTCCTTCGAGCTGATCCGG AGCTCCTTGGCGATGCCATCGACG
At3g12360	32	GCTGAAGTTGCGGAGATTGAGC GTCAAGCGCTGTCTTGTGATCGC
At5g02620	33	CCGGAACCAAAGCCAAGAAGCGC AGCTGAGTATGCACCTCATGGCC
At5g60070	34	AGGGTCAGACGCCACTTCACATG CGTTCTCCGAGAGATTGTCCCG
At5g04690	37	AGAGATTCCGGTTGTGTTGCGG TCCCGTGGACCATAGAAGCTCGG

Los oligonucleótidos específicos diseñados para el análisis de la región genómica en *Arabidopsis* y colza, y para la región transmembrana del gen *AtAnkTm28* se indican a continuación:

Código Atg	EN	Secuencia de los oligonucleótidos (5'a 3')
At4g03450	3	CAGAGAAGTACGAGGGCGTTTCATCTGATC TTCGACACTAGTCGCGACATAAAGATGGACGG
At4g03500	8	GATTTTCCATGGGAAGGCCGGATGTGC TTCATTCCATGGATGCAAGTGTACACATCAACA
At4g14400	11	ATGTCGCCATGGCTCCGGAGATTTTTGGTGGA TGATTTCCATGGAACACGCCACACAACCAA
At1g14480	12	TTTTTACTAGTGATTTGAGACTCCAACAAGCTG TTGTAAACTAGTCCAAAACCAGCTGAATCGGG
At5g15500	16	ACTAATAGATCTAAGATCATTAGAAGCTGC AATTAAGATCTCTAATCACATACCGATTAGGC
At5g54610	18	TTGTAACCATGGATTCAAATTGCTTTTGGT TTCAAACCATGGTCTTGAAGTTGTCAAAAATC
At1g10340	20	TTGAGAAGATCTGCAGCCGATCTTCCATGC GTGAAAAGATCTGATTTTCCGGAGACGGC
At2g24600	22	AAAAAGACTAGTCCGATCTTCGATGCCATC GATTCAACTAGTGAATAGCCTGACCCCTTC
At1g07710	28	TATCAACCATGGAAGGGGAAGAAGACAC CAAGTCCCATGGCATACATTTTCTTACACTCGT ACCCCGCCATGGAAGTGAACAAACCGTAAG
At2g31820	30	GAGCAACCATGGAGAGGCAGAAGAGTTTC CAAATCCCATGGCATACATTCTTATTGTATT
At3g09550	31	AATCTCCCATGGAGAAGAATCTTCTGGATTTGAC TCGATCCCATGGCGTATATCGGATTAACCTCC
At3g12360	32	ATCCTACCATGGATCCTTACCAACACCTTCAC CATCTTCCATGGCGAAAATAGGATCAACTTCC
At5g02620	33	AAGAAGACTAGTACGAAGCAGATGACGGCAAG CTCTAAACTAGTCGCATAAAGTTTCCGGTTATGT

### Genes control

Los genes control utilizados en los análisis de expresión por RT-PCR han sido el gen de la *actina* y el gen *AtEm6* (gen específico del desarrollo de embrión de *Arabidopsis thaliana*).

Para el análisis del patrón de expresión bajo diferentes condiciones de estrés se utilizaron los genes *rab18*, *cor15* y *adh1*.

Código Atg	EN	Secuencia de los oligonucleótidos (5'a 3')
At5g09810	<i>Actina</i>	GGCCGATGGTGAGGATATTC CTGACTCATCGTACTCACTC
At2g40170	<i>AtEm6</i>	GGCGTCTCAACAAGAGAAGAAGC GGGGAAGTTTGATTTAGGTCTTG
At5g66400	<i>rab18</i>	ATGACGAGTACGGAATCCGATGG TATGTATACAGATTGTTCAAGC
At2g42540	<i>cor15</i>	ATGGCGATGTCTTTCTCAGG GCATCCTTAGCCTCTCCTGC
At1g77120	<i>adh1</i>	TCCACGTATCTTCGGCCATG TAGCACCTTCTGCAGCGCC

### Análisis de líneas mutantes de los genes *AtAnkTm* mediante PCR

Los oligonucleótidos fueron diseñados con el fin de poder identificar fácilmente si el fragmento amplificado contenía o no la inserción de T-DNA, o bien si la inserción estaba presente en sólo uno de los alelos. Como control negativo se utilizó DNAg de plantas silvestres de *Arabidopsis*.

LB2, LB3 y QRB3 corresponden a los oligonucleótidos diseñados para la región flanqueante de la inserción.

A, B, C y D, son los correspondientes al gen como se describe en el siguiente cuadro:

Línea SAIL	Gen	Oligonucleótidos para gen sin inserción	Oligonucleótidos para gen con inserción
66	At5g15500	A: AACGCCGACGGACTTACAC B: ATTCCCAAACCAAACCTACC	B: ATTCCCAAACCAAACCTACC LB2: GCTTCCTATTATATCTTCCCAAATTACCAATACA
44	At4g11000	C: GGAAGGTGAGCACTGAGAGATAGC D: CGGAAGTCTCTCTTGGGGCTGAGC	LB2: GCTTCCTATTATATCTTCCCAAATTACCAATACA C: GGAAGGTGAGCACTGAGAGATAGC
20	At5g02620	C: GTCTAAATGACTAAATCCATCGC D: ATCATCTCCAAAAGAAGGTCCG	LB3: GCATCTGAATTCATAACCAATCTCG D: ATCATCTCCAAAAGAAGGTCCG
140	At5g60070	A: AGGGTCAGACGCCACTTCACATG B: CGTTCTCCGAGAGATTGTCCCG	B: CGTTCTCCGAGAGATTGTCCCG QRB3: CAATTCACACAGGAAACAGCTATG
633	At3g09550	A: GTTGCTCCTTCGAGCTGATCCGG B: AGCTCCTTGGCGATGCCATCGACG	A: GTTGCTCCTTCGAGCTGATCCGG LB2: GCTTCCTATTATATCTTCCCAAATTACCAATACA
63	At4g03440	B: CTTGGATTGCTACGTGCGTAGCC C: GAGATACTGCTCTCCAATCAGCC	B: CTTGGATTGCTACGTGCGTAGCC QRB3: CAATTCACACAGGAAACAGCTATG
642	At4g05040	B: CCTGCAGCAAACGTCAATTGTGGC C: GGTGTTGGTCAAGATGTGGATGG	C: GGTGTTGGTCAAGATGTGGATGG LB2: GCTTCCTATTATATCTTCCCAAATTACCAATACA

## 2.8. TRANSFERENCIA E HIBRIDACIÓN DE ÁCIDOS NUCLEICOS

### 2.8.1. Transferencia de ácidos nucleicos

#### 2.8.1.1. Transferencia para Southern

Esta técnica (Southern, 1975) permite la identificación y el análisis de fragmentos de DNA por su fraccionamiento en geles de agarosa, siguiendo por una transferencia a un filtro o membrana. Las membranas utilizadas han sido membranas de nylon (Hybond-N<sup>TM</sup>, Amersham). La movilización de los ácidos nucleicos desde el gel hasta la membrana se consigue por capilaridad, con un tampón adecuado, ascendiendo verticalmente a través de una pila de papel absorbente compactado por presión. Una vez finalizada la transferencia se inmoviliza el DNA a la membrana utilizando el StrataLinker (Stratagene) o bien manteniendo la membrana 1 hora a -80°C. De esta manera, en la membrana ésta representada una copia exacta del patrón de bandas presente en el gel (Brown, 1993).

1. Colocar una placa de vidrio limpia sobre una cubeta que contenga la solución de transferencia.
2. Extender sobre la placa una tira de papel Whatman 3MM empapado en la solución de transferencia que sobresalga de la placa y este en permanente contacto con la solución de transferencia. Debe ser de mayor tamaño que el gel.
3. Colocar uno sobre otro y en el orden siguiente:
  - Gel, cubriendo el resto de la tira de papel Whatman con plástico para evitar el contacto con los papeles absorbentes.
  - Membrana de nailon del mismo tamaño que el gel y tratada según estipule el fabricante.
  - 3 papeles Whatman 3MM del tamaño del gel.
  - Pila de papeles absorbentes de 5-10 cm de grosor.
  - Placa de vidrio.
  - Peso de 0.5-1 Kg.

Es muy importante que no haya burbujas entre el gel y el resto de capas.

4. Dejar transferir un mínimo de 6 horas, siempre es mejor durante toda la noche. El tiempo dependerá de la abundancia de los ácidos nucleicos.
5. Fijar los ácidos nucleicos a la membrana.

### **2.2.1.2. Transferencia para northern**

Mediante esta técnica se pueden transferir por capilaridad mRNAs desde el gel a una membrana de nylon (Hybon-N, Amersham). Para evitar que el RNA se degrade es necesario trabajar con material libre de RNAsas y guantes. Se pueden analizar hasta 30 µg de RNA, aunque entre 10-15 µg son suficientes para detectar RNAs abundantes. En las transferencias Northern se ha utilizado el método del formaldehído, agente desnaturizante.

1. Preparar un gel de agarosa al 1.2%, MEN 1X y formaldehído 20%.
2. Utilizar MEN 1X como tampón de electroforesis.
3. Preparar las muestras de RNA mezclando:
  - 4.7 µl RNA (10-20 µg RNA total y H<sub>2</sub>O libre de RNAsas)
  - 2.0 µl MEN 1X
  - 3.3 µl formaldehído 37%
  - 10 µl formamida desionizada
4. Desnaturizar 15 min a 65°C.
5. Añadir 2 µl de tampón de carga y cargar el gel.
6. Mantener la electroforesis a 70-80 Voltios hasta que el bromofenol haya migrado un 80% de la longitud del gel.
7. Parar la electroforesis y proceder a la transferencia utilizando como solución de transferencia SSC 10X o bien SSPE 20X.
8. Desmontar la transferencia y fijar el RNA a la membrana en el Stratalinker.
9. Teñir la membrana con azul de metileno.
10. Escanear la imagen para comparar los niveles de carga entre las diferentes muestras.
11. La membrana se destiñe totalmente con una solución que contenga SDS 1% y SSC 1%.
12. Almacenar la membrana entre papeles filtro a temperatura ambiente hasta el momento de la utilización.

### **2.8.2. Marcaje y purificación de sonda**

El marcaje de las sondas de DNA utilizadas tanto en las hibridaciones northern ha sido un marcaje radioactivo con [ $\alpha$ -<sup>32</sup>P]dCTP mediante el kit comercial *Random Primed DNA Labeling Kit* de Roche, siguiendo el protocolo establecido por el fabricante con algunas modificaciones. Se recomienda que el fragmento de DNA a marcar tenga entre 100 y 200 pb de longitud y sea lineal. En la purificación de la sonda se han empleado las columnas ProbeQuant™ G-50 *Micro Columns* (Amersham Biosciences) con el fin de eliminar los nucleótidos marcados no incorporados en la sonda.

#### **2.8.2.1. Marcaje**

1. Añadir H<sub>2</sub>O dd a 10 ng-3 µg DNA hasta un volumen final 11 µl.
2. Incubar 10 min a 95°C y transferir a hielo inmediatamente.

3. Añadir:

- 3  $\mu$ l dNTP *stock mix*
- 2  $\mu$ l *reaction mixture*
- 3  $\mu$ l [ $\alpha$ -<sup>32</sup>P]dCTP
- 1  $\mu$ l *Klenow enzyme*

4. Mezclar y centrifugar brevemente.

5. Incubar 1 hora a 37°C

6. Detener la reacción con 2  $\mu$ l EDTA 0.2M (pH 8.0) o calentando 10 min a 65°C.

### 2.8.2.2. Purificación

1. Añadir el volumen necesario de TE pH 8.0 a la reacción de marcaje hasta un volumen final de 50  $\mu$ l, mezclar y centrifugar brevemente.
2. Preparar la columna: Resuspender la resina de la columna vorteadando, abrir levemente el tapón, romper el final de la columna y colocarla en un tubo de microcentrífuga.
3. Añadir la reacción de marcaje a la columna y centrifugar 2 min a 3.000 rpm.
4. Descartar la columna y guardar sonda eluida.

### 2.8.3. Hibridación

Las prehibridaciones, hibridaciones y lavados se han realizado en tubos de metacrilato o vidrio y en estufas rotatorias.

1. Preparar la solución de hibridación/prehibridación y añadir 0.25 mg/ml de DNA de salmón sonicado y desnaturalizado.
2. Poner las membranas en los tubos y lavarlas con tampón fosfato 0.25 M .
3. Prehibridar como mínimo 1 hora en solución prehibridación.
4. Cambiar la solución y añadir la sonda desnaturalizada 5 min a 100°C.
5. Dejar hibridar a la temperatura adecuada mínimo 5 horas y máximo 16 horas.
6. Sacar la sonda y guardarla en un tubo Falcon.
7. Lavar las membranas en las siguientes soluciones:
  - SSC 2X / SDS 0.1%, 2 lavados a 65°C de 20 min.
  - SSC 0.5X / SDS 0.1%, 2 lavados a 65°C de 20 min.
  - SSC 0.1X / SDS 0.1%, 2 lavados a 65°C de 20 min.
  - SSC 0.05X / SDS 0.1%, 2 lavados a 65°C de 20 min

El número de lavados depende de las condiciones de astringencia para la sonda en cada caso.

8. Sellar la membrana dentro de una bolsa de plástico y exponer con pantalla de PhosphorImager de Kodak o bien a -80°C con una pantalla intensificadora.

Normalmente las prehibridaciones e hibridaciones se han realizado a 65°C.

Solución de prehibridación/hibridación: fosfato sódico (Na<sub>2</sub>HPO<sub>4</sub>) 0.25M, SDS 7%, EDTA 1 mM (pH 8.0).

#### **2.8.4. Deshibridación**

La deshibridación de las membranas para posteriores reutilizaciones se ha realizado según el procedimiento descrito por el fabricante (Hybond-N<sup>TM</sup>, Amersham).

1. Verter sobre las membranas SDS 0.1% a punto de ebullición.
2. Dejar atemperar en agitación y repetir nuevamente el tratamiento.
3. Verificar la deshibridación mediante un contador Geiger o exponiendo las membranas.

### **2.9. TÉCNICAS DE DETECCIÓN *IN SITU***

#### **2.9.1. Fijación e inclusión en parafina**

La inclusión en parafina consiste en la sustitución de toda el agua que contienen las muestras por una matriz inerte y nos permite realizar cortes histológicos de 8  $\mu\text{m}$  de grosor. La fijación de las muestras evita la degradación del tejido al detener la actividad enzimática de manera rápida. Así mismo también tiene la finalidad de reforzar el tejido contra los efectos de las posteriores reacciones.

##### **2.9.1.1. Fijación**

1. Cortar los tejidos en pequeños fragmentos de 2 a 4 mm.
2. Tratar los tejidos con solución de fijación formaldehído-EtOH-ácido acético 80:3.5:5 dejar 1 hora a temperatura ambiente.
3. Eliminar el fijador y reemplazar por solución de fijación nueva. Almacenar 1 semana a 4°C.
4. Cambiar la solución por EtOH 70% e incubar 1 hora a temperatura ambiente.
5. Reemplazar por nuevo EtOH 70% y almacenar a 4°C hasta el momento de la inclusión.

##### **2.9.1.2. Inclusión**

1. EtOH 70% 1 hora a Ta.
2. EtOH 80% 1 hora a Ta.
3. EtOH 90% 1 hora a Ta.
4. 0.02% eosina en EtOH 100% 1 hora a Ta y después a 4°C toda la noche.
5. EtOH 100% 1 h hora a Ta.
6. EtOH 100% 1 h hora a Ta.
7. 25% Histo-Clear, 75% EtOH 100% 1 hora a Ta.
8. 50% Histo-Clear, 50% EtOH 100% 1 hora a Ta.
9. 75% Histo-Clear, 25% EtOH 100% 1 hora a Ta.
10. 100% Histo-Clear 1 hora a Ta.
11. 100% Histo-Clear 1 hora a Ta.
12. Vaciar la mitad del vial y rellenar con parafina. Incubar a 60°C toda la noche.



13. Reemplazar con nueva parafina fundida previamente a 60°C e incubar a 60°C.
14. Repetir los cambios tres veces cada 4-6 horas.

### **2.9.1.3. Preparación de los portaobjetos**

Se ha incluido en general una muestra por bloque de parafina utilizando. Los cortes histológicos se efectuaron con microtomo (8 µm). Se utilizaron portaobjetos especiales para la técnica de hibridación *in situ* (Dako™).

### **2.9.1.4. Desparafinación, deshidratación y permeabilización**

El xileno y el TBA han sido sustituidos por HistoClear, una sustancia menos tóxica.

1. Histo-Clear, 10 minutos
2. Histo-Clear, 10 minutos
3. EtOH 100%, 1 minuto
4. EtOH 100%, 30 segundos
5. EtOH 95%, 2 minutos
6. EtOH 80%, 2 minutos
7. EtOH 70%, 2 minutos
8. EtOH 50%, 2 minutos
9. EtOH 30%, 2 minutos
10. H<sub>2</sub>O DEPC, 2 minutos
11. PBS 1X, 5 minutos
12. Tratamiento con Proteinasa K. Precalentar la solución Tris-HCl 100mM (pH 8) y EDTA 50mM a 37°C y añadir 1–2 µg/ml de proteinasa K inmediatamente antes de sumergir los portaobjetos.
13. Incubar a 37°C durante 15 minutos en agitación.
14. Bloquear la reacción incubando 2 minutos en glicina 0.2% en PBS
15. Lavar 2 veces en PBS durante 2 minutos.
16. Incubar en Trietanolamina 0.1M (pH 8) durante 10 minutos.
17. Lavar 2 nuevamente 2 veces en PBS 5 minutos.

### **2.9.2. Síntesis de las ribosondas**

El cDNA debe ser clonado en un vector que contenga los promotores de las RNA polimerasas T7 y SP6, como es el vector pCRII-TOPO (Invitrogen); aunque la eficiencia de la transcripción al usar la RNA polimerasa SP6 no ha sido tan buena. El plásmido debe estar linealizado, antes de iniciar la transcripción, mediante la digestión con enzimas de restricción específicos que permitan obtener las sondas sentido y antisentido del cDNA de interés. Las enzimas deben dejar extremos romos o bien extremos 5' protuberantes. El marcaje de la sonda se realiza por incorporación de UTP–digoxigenina. El inserto no debería tener más de 500 pb, ya que el RNA tiene mayores dificultades para formar el híbrido cuanto más grande sea el inserto. Sin embargo, un inserto demasiado pequeño daría problemas de inespecificidad. Así, la mejor opción es una sonda de entre 200 y 400 pb.

### 2.9.2.1. Linealización

La digestión se realiza con una enzima que corte en una posición que debe estar a continuación del inserto en relación a la posición del promotor de la RNA polimerasa; de esta manera se evita que el plásmido sea también transcrito. Para decidir que transcripción corresponde a la sonda sentido y cual a la sonda antisentido, debe tenerse en cuenta que la transcripción siempre tiene lugar de 5' a 3'; por lo tanto, la cadena copiada es aquella que tiene el sentido contrario, 3' a 5'. Entre 3 y 4  $\mu\text{g}$  de plásmido son suficientes para la reacción. Una vez verificada la digestión en un gel de agarosa 1%, se procede a la extracción con fenol/cloroformo/alcohol isoamílico; la precipitación con 0.1 V de NaAc 3 M (pH 5.5) y 2.5 V de EtOH 100% y el lavado de pellet con EtOH 70%. Finalmente, el DNA es resuspendido en el volumen necesario de H<sub>2</sub>O DEPC para obtener una concentración final de 1  $\mu\text{g}/\mu\text{l}$ .

### 2.9.2.2. Reacción de Transcripción

1. Añadir en un tubo de microcentrífuga los siguientes reactivos:

Tampón de transcripción 5X	4 $\mu\text{l}$
DTT 0.1M	2 $\mu\text{l}$
RNAguard™ (26,8U/ $\mu\text{l}$ )	1 $\mu\text{l}$
DIG RNA <i>labelling mix</i> 10X	2 $\mu\text{l}$
Plásmido (500 ng a 1 $\mu\text{g}$ )	9 $\mu\text{l}$
RNA polimerasa T7, T3 ó SP6	2 $\mu\text{l}$

2. Incubar 2 horas a 37°C.
3. Agregar 1  $\mu\text{l}$  DNAsa RNAsa-free e incubar durante 15 min a 37°C.
4. Verificar la digestión en gel de agarosa 1% con 0.5  $\mu\text{l}$  de RNA.
5. Precipitar con 0.1 V NaAc 3 M (pH 5,2) y 2.5 V de EtOH 100% durante 1 hora a 70 °C o bien o/n a -20°C.
6. Centrifugar 20 min y lavar el pellet con EtOH 70%, centrifugar 20 min y dejar secar a temperatura ambiente.
7. Resuspender en 10  $\mu\text{l}$  de H<sub>2</sub>O libre de RNAsas.

### 2.9.2.3. Cuantificación de las sondas

1. Tomando 1  $\mu\text{l}$  de cada sonda, realizar una serie de diluciones 1/20, 1/250, 1/1000 y 1/2500 y añadirles 91  $\mu\text{l}$  de tampón de hibridación.
2. Depositar 1  $\mu\text{l}$  de cada dilución y para cada sonda en una membrana de nylon (Hybon-N). Fijar la sonda a la membrana mediante Stratalinker.
3. Proceder a la inmunodetección y revelado.
4. Se seleccionará aquellas diluciones más similares entre las dos sondas sentido y antisentido, y más cercana a las de mayor dilución en que se perciba un leve marcaje. (Recordar que las sondas ya están diluidas 1/10 de partida).

### **2.9.3. Hibridación**

Se utilizó entre 100 y 300 µl de tampón de hibridación por cada pareja en sandwich de portaobjetos. Es necesario así, calcular cuanto buffer se usará según el número de portaobjetos para cada sonda.

1. Precalentar los portaobjetos en placa calefactora a 42°C.
2. Mezclar el tampón de hibridación y la cantidad de sonda necesaria en cada caso.
3. Desnaturalizar 2 min a 80°C y transferir a hielo inmediatamente.
4. Repartir 300 µl de solución en cada porta y cubrirlo con otro portaobjetos.
5. Colocarlos en cámara húmeda que contenga SSC 2X y formamida 50%, e incubar O/N a 50°C.
6. Lavar dos veces en solución SSC 2X, formamida 50% durante 90 min.

Tampón de hibridación: *In situ* Sales 1X (pH 6.8), dextran sulfato 50%, Denhardt's 1%, tRNA 1mg/ml, poli A+ 500 µl/ml, formamida 50%.

*In situ* Sales 10X: NaCl 3M, Tris-HCl 100 mM (pH 6.8), Na<sub>2</sub>PO<sub>4</sub> 100 mM (pH 6.8), EDTA 50 mM.

### **2.9.4. Inmunodetección y revelado**

1. Lavar en TBS 1X durante 5 min.
2. Incubar en Solución de bloqueo durante 1 hora.
3. Lavar en BSA 1%, Triton X-100 0.3% en TBS durante 30 min.
4. Incubar con 1:3.000 de anticuerpo AP *conjugated anti-DIG, Fab fragmente* (Roche) diluido en BSA 1% (*Bovine Serum Albumine F-V*), Triton X-100 0.3% durante 90 min.
5. Lavar 3 veces en BSA 1%, TritonX-100 0.3% durante 20 min.
6. Lavar durante 5 min con Tampón de detección.
7. Incubar en oscuridad con 0.8% *NBT/BCIP mix* (Roche) en tampón de detección durante el tiempo que sea necesario hasta observar al microscopio óptico que aparece señal de hibridación.
8. Detener la reacción lavado con H<sub>2</sub>Odd, EtOH 70%, EtOH 100%, EtOH 70% y H<sub>2</sub>Odd, entre 1-5 min.

TBS : Tris-HCl 100 mM (pH 7.5), NaCl 400 mM.

Solución de bloqueo: 0.5% *Boehringer block agent* en TBS 1X.

Tampón de detección: TrisHCl 100 mM (pH 9.5), NaCl 100 mM, MgCl<sub>2</sub> 50 mM.

## **2.10. TINCIONES HISTOLÓGICAS**

### **2.10.1. Tinción DAPI**

El 4',6-Diamidino-2-phenylindole (DAPI) es una sustancia fluorescente que tiñe ácidos nucleicos y nos permite ver los núcleos de las células al observar las muestras en el microscopio de fluorescencia. Para el análisis del polen se siguió el protocolo que se describe a continuación:

1. Extraer las anteras de las flores abiertas y a punto de abrir y colocarlas sobre un portaobjeto;
2. Presionar levemente el cubreobjetos para romper las anteras y liberar el polen;
3. Añadir 150  $\mu$ l de solución DAPI en cada portaobjetos, cubrirlos con cubreobjetos e incubar 30 minutos a 2 horas a temperatura ambiente;
4. Observar en microscopio de fluorescencia,  $\lambda=300-350$  nm.

Solución DAPI (1  $\mu$ g/ml): 0.05 M de NaHPO<sub>4</sub> (pH 7.0), Tritón X-100 0.5%.

### **2.10.2. Tinción Naranja de Acridina**

La tinción con naranja de acridina permite observar y distinguir, gracias al marcaje selectivo de ácidos nucleicos, el DNA y el RNA que contienen las células. Así, en el microscopio de fluorescencia observaremos fluorescencia amarilla para el DNA, naranja para el RNA y en el caso de las células vegetales, la pared celular vegetal con fluorescencia verde.

1. Desparafinado y deshidratación de las muestras como se indica para la técnica de *in situ*.
2. Incubar 5 min en glicina/HCl 0.2M (pH 2.0).
3. Sumergir 30 min en 0.5 mg/ml Naranja de Acridina, en oscuridad.
4. Lavar 2 veces solución glicina/HCl 0.2M (pH 2.0).
5. Añadir medio de montaje Mowiol y cubrir con cubreobjetos.
6. Observar en microscopio de fluorescencia,  $\lambda= 300-350$  nm.

### **2.10.3. Tinción Azul de Anilina**

El azul de anilina contiene un fluorocromo (sirofluor) que se une específicamente a  $\beta$ -1,3-glucano, el principal componente de la pared del tubo del polen. La preparación de las muestras se describe a continuación:

1. Extraer las anteras de las flores abiertas y a punto de abrir y colocarlas sobre un portaobjeto.
2. Presionar levemente el cubreobjetos para romper las anteras y liberar el polen.
3. Añadir 150  $\mu$ l de solución de azul de anilina en cada portaobjetos, cubrirlos con cubreobjetos e incubar 30 minutos a 2 horas a temperatura ambiente.
4. Observar en microscopio de fluorescencia,  $\lambda=300-350$  nm.

Solución de azul de anilina: 0,1% de azul de anilina (w/v).

## **2.11. TRANSFERENCIA DE DNA MEDIANTE MICROBOMBARDEO**

Se ha empleado el aparato PDS1000/He de BioRad siguiendo el método de Klein (1988) con algunas modificaciones.

### **2.11.1. Preparación de los microproyectiles**

1. Lavar las partículas de oro (1 $\mu$ m diámetro) con EtOH absoluto (HPLC), agitando en vórtex 10 min.
2. Centrifugar 1min a 10.000 rpm y eliminar el sobrenadante.
3. Lavar con Glicerol 50% estéril y vortear 30 s a máxima velocidad.
4. Centrifugar 1 min a 10.000 rpm y eliminar el sobrenadante.
5. Repetir el lavado 2 veces más.
6. Resuspender las partículas en 1 ml de glicerol 50% estéril
7. Almacenar a -20°C.

### **2.11.2. Precipitación del DNA**

1. Añadir secuencialmente y agitar a máxima velocidad 5 min después de cada paso:
  - 10  $\mu$ l DNA (0.5  $\mu$ g/ $\mu$ l)
  - 100  $\mu$ l H<sub>2</sub>O estéril
  - 38  $\mu$ l partículas de oro (previamente vorteadas 5 min)
  - 150  $\mu$ l de CaCl<sub>2</sub>/espermidina (125  $\mu$ l CaCl<sub>2</sub> y 25  $\mu$ l espermidina 0.1M)
2. Sedimentar las partículas en hielo en posición vertical 15 min.
3. Descartar el sobrenadante y añadir 500  $\mu$ l EtOH absoluto (HPLC) y vortear 20 s.
4. Repetir los pasos 2 y 3.
5. Resuspender en 15  $\mu$ l EtOH absoluto (HPLC), sonicar 2 min y vortear 10 min.
6. Dividir en 2 (para los dos disparos por muestra) y mantener en hielo.

### **2.11.3. Preparación PDS1000/He y Bombardeo**

1. Encender la cabina de flujo laminar. Esterilizar el aparato y las piezas con EtOH absoluto. En placas Petri, esterilizar previamente con EtOH absoluto, las membranas portadoras y las rejillas de parada. Es muy importante que esté todo bien seco para su correcta utilización.
2. Encender la bomba de vacío. Abrir el grifo de la bomba de helio y ajustar a 200 psi (*pound per square inch*) por encima de la presión de resistencia de los discos de rotura.
3. Ajustar correctamente la membrana portadora sobre el soporte que debe estar en una superficie horizontal (comprobar con el nivel). Añadir 8  $\mu$ l de EtOH absoluto en el centro de la membrana y dejar secar. Preparar dos membranas por muestra para los dos disparos.
4. Depositar 8  $\mu$ l de muestra (previamente vorteadas) sobre la membrana portadora sin tocarla y dejar secar 5 min.
5. Poner una rejilla de parada en el soporte disparador y a continuación el soporte con la membrana portadora de manera que la superficie con la muestra mire la rejilla. La distancia entre ambas debe ser 11 mm.
6. Poner el disco de ruptura en su soporte y enroscarlo en el extremo del cilindro de gas hasta que quede fijo. Ajustar con la palanquita.
7. Colocar el soporte disparador ya montado en el nivel 5 (el más cercano a la base). La distancia entre éste y el soporte de ruptura ha de ser 11 mm.

8. Poner la placa (sin la tapa) con la muestra vegetal en el nivel 3 y cerrar la cámara.
9. Accionar la bomba de vacío hasta que el manómetro llegue a 27 (0,1 atm). Accionar el botón de disparo hasta escuchar la ruptura del disco.
10. Abrir el vacío inmediatamente y esperar que suba de nuevo la presión a 27. Entonces sacar la muestra.
11. Sacar el soporte de ruptura y comprobar que el disco se ha roto.
12. Sellar la placa, envolverlas en papel de aluminio y dejarlas de 18 a 48 h a 28 °C antes de mirar al microscopio.

## 2.12. MICROSCOPIA ELECTRÓNICA DE BARRIDO

El polen de las líneas mutantes letales en homocigosis y el control Col-0 fueron sometidos a microscopía electrónica de barrido. Para ello se utilizó el método de la acetólisis (Erdtman, 1960) para eliminar el contenido interior y dejar sólo la cubierta (exina) del polen (Ver Introducción, apartado 1.5). El protocolo de preparación de muestras se describe a continuación:

1. Tomar al menos 10–15 flores en similar estado fenológico (previo a la apertura completa de la flor).
2. Extraer los estambres y colocarlos sobre un papel filtro especial de 1 cm de diámetro.
3. Agregar 1 gota de etanol 96% para liberar el polen de la antera.
4. Dejar secar y retirar cuidadosamente los restos de anteras que están repartidos encima del papel filtro.
5. Preparar en el momento ácido acético naciente que consiste en: anhídrido acético tratado con ácido sulfúrico concentrado (9:1). Esta es una reacción exotérmica y debe ser realizada bajo campana de gases.
6. Agregar 1 gota de ácido acético naciente sobre el filtro que contiene los granos de polen.
7. Dejar secar en cabina de flujo.
8. (Opcional) Se puede agregar etanol 96% para eliminar los restos del contenido del polen que están sobre el filtro.
9. Dejar secar 16 – 20 horas en campana de gases. Las muestras pueden ser recubiertas con oro una vez completado este período.

El recubrimiento con oro y la observación en microscopio electrónico de barrido (Hitachi S-2300) se llevaron a cabo utilizando las instalaciones de los servicios científico-técnicos de la Universidad de Barcelona. El recubrimiento con oro fue realizado por los técnicos del servicio. Las imágenes fueron almacenadas utilizando el programa Quartz PCI.

## 2.13. ANÁLISIS *IN SILICO*

Los diversos programas informáticos y bases de datos empleados en la realización de esta memoria se indican a continuación:

- *Gene Expression Visualization de AtGenExpress:*  
<http://www.weigelworld.org/resources/microarray/AtGenExpress>

- Northern Digital y Meta-Analyzer de *Genevestigator*<sup>®</sup>  
<https://www.genevestigator.ethz.ch/>
- *Expressed Sequences Tags* (ESTs) de la base de datos de Arabidopsis en TIGR (The Institute for Genomic Research)  
[http://www.tigr.org/tigr-scripts/tgi/T\\_index.cgi?species=arab](http://www.tigr.org/tigr-scripts/tgi/T_index.cgi?species=arab)
- Bases de datos de líneas mutantes de inserción:
  - NASC (The Nottingham Arabidopsis Stock Centre) y ABRC (Arabidopsis Biological Resource Center). <http://arabidopsis.info/>
  - SIGnAL (The Salk Institute Genome Analysis Laboratory)  
<http://signal.salk.edu/cgi-bin/tdnaexpress>
  - Base de datos de inserciones. <http://atidb.org/>
- Seedgenes: base de datos de genes de Arabidopsis implicados en la embriogénesis. <http://www.seedgenes.org/>.
- dbEST: base de ESTs en GeneBank:  
<http://www.ncbi.nlm.nih.gov/dbEST/index.html/>
- Micromatrices de Affymetrix. <http://affymetrix.arabidopsis.info/>
- Base de datos sobre Arabidopsis. [www.arabidopsis.org](http://www.arabidopsis.org)
- SMART v3.5 : determinación de dominios proteícos:  
<http://smart.embl-heidelberg.de/>
- REP v1.1: determinación de repeticiones en proteínas:  
<http://www.embl-heidelberg.de/~andrade/papers/rep/search.html>
- BLAST: búsqueda de secuencias similares en los bancos:  
<http://www.ncbi.nlm.nih.gov/BLAST/>
- CLUSTALW: alineamientos de secuencias. <http://www.ebi.ac.uk/clustalw/>
- Análisis de dominios transmembrana: TMHMM v. 2.0:  
<http://www.cbs.dtu.dk/services/TMHMM/>
- Análisis SOTA (*Self-organizing tree algorithm*) en TMEV (TIGR Multiple Experiment Viewer): <http://www.tm4.org/mev.html/>

## **BIBLIOGRAFÍA**



- Aarts M, Corzaan P, Stiekema W, Pereira A. 1995. A two-element Enhancer-Inhibitor transposon system in *Arabidopsis thaliana*. *Mol Gen Genet* 247: 555-564.
- Adam D. 2000. *Arabidopsis thaliana* genome. Now for the hard ones. *Nature* 408: 792-793.
- Adams MD, Celniker SE, Holt RA, Evans CA, Gocayne JD, Amanatides PG, Scherer SE, Li PW, Hoskins RA, Galle RF, George RA, Lewis SE, Richards S, Ashburner M, Henderson SN, Sutton GG, Wortman JR, Yandell MD, Zhang Q, Chen LX, Brandon RC, Rogers YH, Blazej RG, Champe M, Pfeiffer BD, Wan KH, Doyle C, Baxter EG, Helt G, Nelson CR, Gabor GL, Abril JF, Agbayani A, An HJ, Andrews-Pfannkoch C, Baldwin D, Ballew RM, Basu A, Baxendale J, Bayraktaroglu L, Beasley EM, Beeson KY, Benos PV, Berman BP, Bhandari D, Bolshakov S, Borkova D, Botchan MR, Bouck J, Brokstein P, Brottier P, Burtis KC, Busam DA, Butler H, Cadieu E, Center A, Chandra I, Cherry JM, Cawley S, Dahlke C, Davenport LB, Davies P, de Pablos B, Delcher A, Deng Z, Mays AD, Dew I, Dietz SM, Dodson K, Doup LE, Downes M, Dugan-Rocha S, Dunkov BC, Dunn P, Durbin KJ, Evangelista CC, Ferraz C, Ferriera S, Fleischmann W, Fosler C, Gabrielian AE, Garg NS, Gelbart WM, Glasser K, Glodek A, Gong F, Gorrell JH, Gu Z, Guan P, Harris M, Harris NL, Harvey D, Heiman TJ, Hernandez JR, Houck J, Hostin D, Houston KA, Howland TJ, Wei MH, Ibegwam C, Jalali M, Kalush F, Karpen GH, Ke Z, Kennison JA, Ketchum KA, Kimmel BE, Kodira CD, Kraft C, Kravitz S, Kulp D, Lai Z, Lasko P, Lei Y, Levitsky AA, Li J, Li Z, Liang Y, Lin X, Liu X, Mattei B, McIntosh TC, McLeod MP, McPherson D, Merkulov G, Milshina NV, Mobarry C, Morris J, Moshrefi A, Mount SM, Moy M, Murphy B, Murphy L, Muzny DM, Nelson DL, Nelson DR, Nelson KA, Nixon K, Nusskern DR, Pacleb JM, Palazzolo M, Pittman GS, Pan S, Pollard J, Puri V, Reese MG, Reinert K, Remington K, Saunders RD, Scheeler F, Shen H, Shue BC, Siden-Kiamos I, Simpson M, Skupski MP, Smith T, Spier E, Spradling AC, Stapleton M, Strong R, Sun E, Svirskas R, Tector C, Turner R, Venter E, Wang AH, Wang X, Wang ZY, Wassarman DA, Weinstock GM, Weissenbach J, Williams SM, Woodage T, Worley KC, Wu D, Yang S, Yao QA, Ye J, Yeh RF, Zaveri JS, Zhan M, Zhang G, Zhao Q, Zheng L, Zheng XH, Zhong FN, Zhong W, Zhou X, Zhu S, Zhu X, Smith HO, Gibbs RA, Myers EW, Rubin GM, Venter JC. 2000. The genome sequence of *Drosophila melanogaster*. *Science* 287: 2185-2195.
- Albert S, Després B, Guilleminot J, Bechtold N, Pelletier G, Delseny M, Devic M. 1999. The EMB506 gene encodes a novel ankyrin repeat containing protein that is essential for the normal development of *Arabidopsis* embryos. *Plant J* 17: 169-179.
- Andrade M, Ponting C, Gibson T, Bork P. 2000. Homology-based method for identification of protein repeats using statistical significance estimates. *J Mol Biol* 298: 521-537.
- Apic G, Gough J, Teichmann SA. 2001. Domain combinations in archaeal, eubacterial and eukaryotic proteomes. *J Mol Biol* 310: 311-325.
- Arevalo JC, Yano H, Teng KK, Chao MV. 2004. A unique pathway for sustained neurotrophin signaling through an ankyrin-rich membrane-spanning protein. *EMBO J* 23: 2358-2368.
- Arniges M, Fernandez-Fernandez JM, Albrecht N, Schaefer M, Valverde MA. 2006. Human TRPV4 Channel Splice Variants Revealed a Key Role of Ankyrin Domains in Multimerization and Trafficking. *J Biol Chem* 281: 1580-1586.
- Azpiroz-Leehan R, Feldmann K. 1997. T-DNA insertion mutagenesis in *Arabidopsis* going back and forth. *Trends Genet* 13: 152-156.
- Bai S, Chen L, Yund MA, Sung ZR. 2000. Mechanisms of plant embryo development. *Curr Top Dev Biol* 50: 61-88.
- Bancroft I, Bhatt A, Sjodin C, Scofield S, Jones J, Dean C. 1992. Development of an efficient two-element transposon tagging system in *Arabidopsis thaliana*. *Molecular & general genetics* 233 (3): 449-461.
- Baranova AV, Lobashev AV, Ivanov DV, Krukovskaya LL, Yankovsky NK, Kozlov AP. 2001. *In silico* screening for tumour-specific expressed sequences in human genome. *FEBS Lett* 508: 143-148.
- Barceló J, Nicolás G, Sabater B, Sánchez R. 2003. Fisiología vegetal. Colección Ciencia y Técnica, Ediciones Pirámide, Madrid.

- Bardwell VJ, Treisman R. 1994. The POZ domain: a conserved protein-protein interaction motif. *Genes Dev* 8: 1664–1677.
- Bateman A. 1997. The structure of a domain common to archaeobacteria and the homocystinuria disease protein. *Trends Biochem Sci* 22: 12-13.
- Becerra C, Jahrmann T, Puigdomenech P, Vicient CM. 2004. Ankyrin repeat-containing proteins in Arabidopsis: characterization of a novel and abundant group of genes coding ankyrin-transmembrane proteins. *Gene* 340: 111-121.
- Becerra C, Puigdomènech P, Vicient CM. 2006. Computational and experimental analysis identifies Arabidopsis genes specifically expressed during early seed development. *BMC Genomics*, 7: 38.
- Bennet V, Baines A. 2001. Spectrin and ankyrin-based pathways: metazoan inventions for integrating cells into tissues. *Physiol Rev* 81: 1353-1392.
- Bennett MD, Leitch IJ, Price HJ, Johnston JS. 2003. Comparisons with Caenorhabditis (100 Mb) and Drosophila (175 Mb) using flow cytometry show genome size in Arabidopsis to be 157 Mb and thus 25 % larger than the Arabidopsis genome initiative estimate of 125 Mb. *Ann Bot* 91: 547–557.
- Bennett V. 1992. Ankyrins: adaptors between diverse plasma membrane proteins and the cytoplasm. *J. Biol Chem* 267: 8703–8706.
- Bennetzen J. 2002. The rice genome. Opening the door to comparative plant biology. *Science* 296: 60-63.
- Bent A. 2000. Arabidopsis in planta transformation. Uses, mechanisms, and prospects for transformations of other species. *Plant Physiol* 124: 1540-1547.
- Berardini TZ, Mundodi S, Reiser L, Huala E, Garcia-Hernandez M, Zhang P, Mueller LA, Yoon J, Doyle A, Lander G, Moseyko N, Yoo D, Xu I, Zoeckler B, Montoya M, Miller N, Weems D, Rhee SY. 2004. Functional annotation of the Arabidopsis genome using controlled vocabularies. *Plant Physiol* 135: 745-755.
- Berleth T, Jürgens G. 1993. The role of the *monopteros* gene in organizing the basal body region of the Arabidopsis embryo. *Development* 118: 575-587.
- Berleth T. 1998. Experimental approaches to Arabidopsis embryogenesis. *Plant Phys Bioch* 36: 69-82.
- Bernstein SL, Borst DE, Neuder ME, Wong P. 1996. Characterization of the human fovea cDNA library and regional differential gene expression in the human retina. *Genomics* 32: 301-308.
- Bevan M, Mayer K, White O, Eisen J, Preuss D, Bureau T, Salzberg S, Mewes H. 2001. Sequence and analysis of the Arabidopsis genome. *Current Opin Plant Biol* 4: 105-110.
- Bhatt AM, Canales C & Dickinson HG 2001 Plant meiosis: the means to 1N. *Trends Plant Sci* 6: 114–121.
- Bolker J. 1995. Model systems in developmental biology. *Bioessays* 17: 451-455.
- Borisjuk L, Rolletschek H, Radchuk R, Weschke W, Wobus U, Weber H. 2004. Seed development and differentiation: a role for metabolic regulation. *Plant Biol* 6: 375-386.
- Bork P. 1993. Hundreds of ankyrin-like repeats in functionally diverse proteins: mobile modules that cross phyla horizontally?. *Proteins* 17: 363–374.
- Bortoluzzi S, d'Alessi F, Romualdi C, Danieli GA. 2000. The human adult skeletal muscle transcriptional profile reconstructed by a novel computational approach. *Genome Res* 10: 344-349.
- Bouche N, Bouchez D. 2001. Arabidopsis gene knockout: phenotypes wanted. *Curr Opin Plant Biol* 4: 111-117.
- Bouché N, Scharlat A, Snedden W, Bouchez D, Fromm H. 2002. A novel family of calmodulin-binding transcription activators in multicellular organisms. *The J Biol Chem* 277: 21851-21861.

- Breeden L, Nasmyth K. 1987. Similarity between cell-cycle genes of budding yeast and fission yeast and the *Notch* gene of *Drosophila*. *Nature* 329: 651-654.
- Brown T. 1993. Analysis of DNA sequences by blotting and hybridization. In: *Current Protocols in Molecular Biology*. Volume 1 (U.S., Wiley Interscience), pp. 2.9.1-2.9.15.
- Bruggemann E, Handwerger K, Essex C, Storz G. 1996. Analysis of fast neutron-generated mutants at the *Arabidopsis thaliana* *HY4* locus. *Plant J* 10: 755-760.
- Cao H, Glazebrook J, Clarke J, Volko S, Dong X. 1997. The *Arabidopsis* NPR1 gene that controls systemic acquired resistance encodes a novel protein containing ankyrin repeats. *Cell* 88: 57-63.
- Cao H, Li X, Dong X. 1998. Generation of broad-spectrum disease resistance by overexpression of an essential regulatory gene in systemic acquired resistance. *PNAS USA* 95: 6531-6536.
- Chang S, Low P. 2003. Identification of a critical ankyrin-binding loop on the cytoplasmic domain on erythrocyte membrane band 3 by crystal structure analysis and site-directed mutagenesis. *J Biol Chem* 278: 6879-6884.
- Chaudhury AM, Koltunow A, Payne T, Luo M, Tucker MR, Dennis ES, Peacock WJ. 2001. Control of early seed development. *Annu Rev Cell Dev Biol* 17: 677-699.
- Chen JJ, Lee S, Zhou G, Rowley JD, Wang SM. 2003. Generation of longer cDNA fragments from SAGE tags for gene identification. *Meth Mol Biol* 221: 207-222.
- Chinchilla D, Merchan F, Megias M, Kondorosi A, Sousa C, Crespi M. 2003. Ankyrin protein kinases: a novel type of plants kinase gene whose expression is induced by osmotic stress in alfalfa. *Plan Molecular Biology*, 51: 555-566.
- Choi Y, Gehring M, Johnson L, Hannon M, Harada J, Goldberg R, Jacobsen S, Fischer R. 2002. DEMETER, a DNA glycosylase domain protein, is required for endosperm gene imprinting and seed viability in *Arabidopsis*. *Cell* 110: 33-42.
- Chye M, Li H, Yung M. 2000. Single amino acid substitutions at the acyl-CoA-binding domain interrupt 14[C]palmitoyl-CoA binding of ACBP2, an *Arabidopsis* acyl-CoA-binding protein with ankyrin repeats. *Plant Mol Biol* 44: 711-721.
- da Costa e Silva O. 1994. CG-1, a parsley light-induced DNA-binding protein. *Plant Mol Biol*. 25: 921-924.
- Daram P, Urbach S, Gaymard F, Sentenac H, Chérel I. 1997. Tetramerization of the AKT1 plant potassium channel involves its C-terminal cytoplasmic domain. *EMBO journal* 16: 3455-3463.
- Delseny M, Bies-Etheve N, Carles C, Hull G, Vicent CM, Raynal M, Grellet F, Aspart L. 2001. Late Embryogenesis Abundant (LEA) protein gene regulation during *Arabidopsis* seed maturation. *J Plant Phys* 158: 419-427.
- Devi S, Binz K, Stumpp M, Plückthun A, Bosshard H, Jelesarov, I. 2004. Folding of a designed simple ankyrin repeat protein. *Protein Sci* 13: 2864-2870.
- Doebley J, Lukens L. 1998. Transcriptional regulators and the evolution of plant form. *Plant Cell* 10: 1075-1082.
- Dolan L, Janmaat K, Willemsen P, Linstead S, Poethig K, Roberts K, Scheres B. 1993. Cellular organization of the *Arabidopsis thaliana* root. *Development* 119: 71-84.
- Dolferus R, Jacobs M, Peacock WJ, Dennis ES. 1994. Differential interactions of promoter elements in stress responses of the *Arabidopsis* *Adh* gene. *Plant Physiol* 105: 1075-1087.
- Dong J, Keller WA, Yan W, Georges F. 2004. Gene expression at early stages of *Brassica napus* seed development as revealed by transcript profiling of seed-abundant cDNAs. *Planta* 218: 483-491.
- Ehrhardt T, Zimmermann S, Müller-Röber B. 1997. Association of plant  $K_{in}^+$  channels is mediated by conserved C-termini and does not affect subunit assembly. *FEBS Lett* 409: 166-170.

- Ekman DR, Lorenz WW, Przybyla AE, Wolfe NL, Dean JF. 2003. SAGE analysis of transcriptome responses in *Arabidopsis* roots exposed to 2,4,6-trinitrotoluene. *Plant Physiol* 133: 1397-406.
- Elster R, Bommert P, Sheridan W, Werr W. 2000. Analysis of four embryo-specific mutants in *Zea mays* reveals that incomplete radial organization of the proembryo interferes with subsequent development. *Dev Genes Evol* 11: 938-943.
- Engel ML, Chaboud A, Dumas C, McCormick S. 2003. Sperm cells of *Zea mays* have a complex complement of mRNAs. *Plant J* 34:697-707.
- Erdtman G (1960) The acetolysis method. *Sven Bot Tidskr* 54: 561-564.
- Fahn A. 1982. *Plant anatomy*. Pergamon Press, New York.
- Fedorova M, van de Mortel J, Matsumoto PA, Cho J, Town CD, VandenBosch KA, Gantt JS, Vance CP. 2002. Genome-wide identification of nodule-specific transcripts in the model legume *Medicago truncatula*. *Plant Physiol* 130: 519-537.
- Fizames C, Munos S, Cazettes C, Nacry P, Boucherez J, Gaymard F, Piquemal D, Delorme V, Combes T, Doumas P, Cooke R, Marti J, Sentenac H, Gojon A. 2004. The *Arabidopsis* root transcriptome by serial analysis of gene expression. Gene identification using the genome sequence. *Plant Physiol* 134: 67-80.
- Gatehouse J, Evans I, Croy R, Boulter, D. 1986. Differential expression of genes during legume seed development. *Philos Transac Royal Soc London. Series B, Biol Sci* 314: 367-384.
- Gavarayeba N. 1996. Sporoderm development in *Liriodendron chinense* (*Magnoliaceae*): a probable role of the endoplasmic reticulum. *Nordic J Bot* 16: 307-323.
- Gerstein M. 1988. How representative are the known structures of the proteins in a complete genome? A comprehensive structural census. *Fold Des* 3: 497-512.
- Giraudat J, Hauge BM, Valon C, Smalle J, Parcy F, Goodman HM. 1992. Isolation of the *Arabidopsis* ABI3 gene by positional cloning. *Plant Cell* 4: 1251-1261.
- Givskov M, Olsen L, Moli S. 1988. Cloning and expression in *E. coli* of the gene for extracellular phospholipase A1 from *Serratia liquefaciens*. *J Bact* 170: 5855-5862.
- Goff SA, Ricke D, Lan TH, Presting G, Wang R, Dunn M, Glazebrook J, Sessions A, Oeller P, Varma H, Hadley D, Hutchison D, Martin C, Katagiri F, Lange BM, Moughamer T, Xia Y, Budworth P, Zhong J, Miguel T, Paszkowski U, Zhang S, Colbert M, Sun WL, Chen L, Cooper B, Park S, Wood TC, Mao L, Quail P, Wing R, Dean R, Yu Y, Zharkikh A, Shen R, Sahasrabudhe S, Thomas A, Cannings R, Gutin A, Pruss D, Reid J, Tavtigian S, Mitchell J, Eldredge G, Scholl T, Miller RM, Bhatnagar S, Adey N, Rubano T, Tusneem N, Robinson R, Feldhaus J, Macalma T, Oliphant A, Briggs S. 2002. A draft sequence of the rice genome (*Oryza sativa* L. ssp. *japonica*). *Science*. 296 (5565): 92-100.
- Goffeau A, Barrell B, Bussey H, Davis R, Dujon B, Feldmann H, Galibert F, Hoheisel J, Jacq C, Johnston M, Louis E, Mewes H, Murakami Y, Philippsen P, Tettelin H, Oliver S. 1996. Life with 6000 genes. *Science* 274: 563-567.
- Goldberg R, de Paiva R, Yadegari R. 1994. Plant embryogenesis: zygote to seed. *Science* 266: 605-614.
- Haberer G, Hindemitt T, Meyers BC, Mayer KF. 2004. Transcriptional similarities, dissimilarities, and conservation of cis-elements in duplicated genes of *Arabidopsis*. *Plant Physiol* 136: 3009-3022.
- Hall A, Fiebig A, Preuss D. 2002. Beyond the *Arabidopsis* genome: opportunities for comparative genomics. *Plant Phys* 129: 1439-1447.
- Hamann T, Mayer U, Jürgens G. 1999. The auxin-insensitive *bodenlos* mutations affects primary root formation and apical-basal patterning in the *Arabidopsis* embryo. *Development* 126: 1387-1395.
- Harada J. 1999. Signaling in plant embryogenesis. *Curr Op Plant Biol* 2: 23-27.

- Haskill S, Beg A, Tompkins S, Morris J, Yurochko A, Sampson-Johannes A, Mondal K, Ralph P, Baldwin A. 1991. Characterization of an immediate-early gene induced in adherent monocytes that encodes I-kappa-B-like activity. *Cell* 65: 1281-1289.
- Haslekas C, Stacy RA, Nygaard V, Culiñez-Macia FA, Aalen RB. 1998. The expression of a peroxiredoxin antioxidant gene, *AtPer1*, in *Arabidopsis thaliana* is seed-specific and related to dormancy. *Plant Mol Biol* 36: 833-845.
- Haughn G, Chaudhury A. 2005. Genetic analysis of seed coat development in *Arabidopsis*. *Trends Plant Sci* 10: 472-427.
- Hemerly A, de Almeida J, Bergonioux C, van Montagu M, Engler G, Inzé D, Ferreira P. 1995. Dominant negative mutants of the Cdc2 kinase uncouple cell division from iterative plant development. *EMBO J* 14: 3925-3936.
- Hemerly A, Ferreira P, Van Montagu M, Engler G, Inzé D. 2000. Cell divisions events are essential for embryo patterning and morphogenesis: studies on dominant-negative *cdc2aAt* mutants of *Arabidopsis*. *Plant J* 23: 123-130.
- Hemerly A, Ferreira P, Van Montagu M, Inzé D. 1999. Cell cycle control and plant morphogenesis: is there an essential link. *BioEssays* 21: 29-37.
- Hony D, Twell D. 2004. Transcriptome analysis of haploid male gametophyte development in *Arabidopsis*. *Genome Biol* 5: R85.
- Hope I. 1994. *Caenorhabditis elegans*, the nematode worm. In J. Bard (ed.), *Embryos. Color Atlas of Development*, Mosby-Year Book Europe, London, pp. 55-75.
- Hudson A. 2000. Development of symmetry in plants. *Ann Rev Plant Phys Plant Mol Biol* 51: 349-370.
- Huminięcki L, Bicknell R. 2000. *In silico* cloning of novel endothelial-specific genes. *Genome Res* 10: 1796-1806.
- Imin N, Kerim T, Weinman JJ, Rolfe BG. 2001. Characterisation of rice anther proteins expressed at the young microspore stage. *Proteomics* 1: 1149-1161.
- International Human Genome Consortium. 2001. Initial sequencing and analysis of the human genome. *Nature* 409: 860-921.
- Isono K, Yamamoto H, Sato K, Kobayashi H. 1999. An *Arabidopsis* cDNA encoding a DNA-binding protein that is highly similar to the DEAH family of RNA/DNA helicase genes. *Nucl Acids Res* 27: 3728-3735.
- Itoh K, Okubo K, Utiyama H, Hirano T, Yoshii J, Matsubara K. 1998. Expression profile of active genes in granulocytes. *Blood* 15: 1432-1441.
- Jahrman T, Bastida M, Pineda M, Gasol E, Ludevid D, Palacín M, Puigdomènech P. 2005. Studies on the function of TM20, a transmembrane protein present in cereal embryos. *Planta* 222: 80-90.
- Jahrman T. 2002. Functional studies on the transmembrane protein encoded by the TM20 gene in maize. Ph. D. Thesis, Univ. Barcelona.
- Jebanathirajah JA, Peri S, Pandley A. 2002. Toll and interleukin-1 receptor (TIR) domain-containing proteins in plants: a genomic perspective. *Trends Plant Sci* 7: 388-391.
- Jonas-Straube E, Hutin C, Hoffman N., Schünemann D. 2001. Functional analysis of the protein-interacting domains of chloroplast SRP43. *J Biol Chem* 276: 24654-24660.
- Jose-Estanyol M, Ruiz-Avila L, Puigdomenech P. 1992. A maize embryo-specific gene encodes a proline-rich and hydrophobic protein. *Plant Cell* 4: 413-423.
- Jung SH, Lee JY, Lee DH. 2003. Use of SAGE technology to reveal changes in gene expression in *Arabidopsis* leaves undergoing cold stress. *Plant Mol Biol* 52: 553-567.
- Jürgens G, Mayer U. 1994. *Arabidopsis*. En: *Embryos*, Bard, J. (ed.), Wolfe Publishing, London, pp. 7-21.

- Jürgens G, Torres-Ruiz R, Berleth, T.. 1994. Embryonic pattern formation in flowering plants. *Ann Rev Gen* 28: 351-371.
- Jürgens G. 1992. Genes to greens: embryonic pattern formation in plants. *Science* 256: 487-488.
- Jürgens G. 1995. Axis formation in plant embryogenesis: cues and clues. *Cell* 81: 467-470.
- Jürgens G. 2001. Apical-basal pattern formation in *Arabidopsis* embryogenesis. *EMBO J* 20: 3609-3616.
- Jürgens, G. Mayer U, Torres-Ruiz R, Berleth T, Miséra S. 1991. Genetic analysis of pattern formation in the *Arabidopsis* embryo. *Development (Suppl.1)*: 27-38.
- Ketchum K, Slayman C. 1996. Isolation of an ion channel gene from *Arabidopsis thaliana* using the H5 signature sequence from voltage-dependent K<sup>+</sup> channels. *FEBS Let* 378 : 19-26.
- Kiyatkin N, Dulubova I, Grishin E. 1993. Cloning and structural analysis of alpha-latroinsectotoxin cDNA. *Eur J Bioch* 213: 121-127.
- Klein TM, Wolf ED, Wu R, Sanford JC. 1987. High-velocity microprojectiles for delivering nucleic acids into living cells. *Nature* 327: 70-73.
- Klimyuk V, Persello-Cartieaux F, Havaux M, Contard-David P, Schuenemann D, Meierhoff K, Gouet P, Jones J, Hoffman, N., Nussaume L. 1999. A chromodomain protein encoded by the *Arabidopsis* CAO gene is a plant-specific component of the chloroplast signal recognition particle pathway that is involved in LHCP targeting. *Plant Cell* 11: 87-99.
- Kohl A, Binz K, Forrer P, Stumpp M, Plückthum A, Grütter M. 2003. Designed to be stable: Crystal structure of a consensus ankyrin repeat protein. *PNAS USA*. 100: 1700-1705.
- Koonin EV, Wolf YI, Karev GP. 2002. The structure of the protein universe and genome evolution. *Nature* 420: 218-223.
- Kuhlmann M, Horvay K, Strathmann A, Heinekamp, T., Fischer U, Böttner S, Dröge-Laser W. 2003. The  $\alpha$ -helical D1 domain of the tobacco b-ZIP transcription factor BZI-1 interacts with the ankyrin-repeats protein ANK1 and is important for BZI-1 function, both in auxin signaling and pathogen response. *J Biol Chem* 278: 8786-8794.
- Kuromori T, Hirayama T, Kiyosue Y, Takabe H, Mizukado S, Sakurai T, Akiyama K, Kamiya A, Ito T, Shinozaki K. 2004. A collection of 11 800 single-copy Ds transposon insertion lines in *Arabidopsis*. *Plant J* 37: 897-905.
- LaMarco K, Thompson C, Byers B, Walton E, McKnight S. 1991. Identification of Ets- and notch-related subunits in GA binding protein. *Science* 253: 789-792.
- Larsson P, Claesson H, Kennedy B. 1998. Multiple splice variants of the human calcium-independent phospholipase A<sub>2</sub> and their effect on enzyme activity. *J Biol Chem* 273: 207-214.
- Laux T, Jürgens G. 1997. Embryogenesis: a new start in life. *Plant Cell* 9: 989-1000.
- Laux, T., Wurschum, T., Breuninger H. 2004. Genetic regulation of embryonic pattern formation. *Plant cell* 16, Suppl.: S190-S202.
- Le Gall M, Giniger E. 2004. Identification of two binding regions for the suppressor of hairless protein within the intracellular domain of *Drosophila* notch. *J Biol Chem* 279: 29418-29426.
- Lee JY, Lee DH. 2003. Use of serial analysis of gene expression technology to reveal changes in gene expression in *Arabidopsis* pollen undergoing cold stress. *Plant Physiol* 132: 517-529.
- Lehti-Shiu MD, Adamczyk BJ, Fernandez DE. 2005. Expression of MADS-box genes during the embryonic phase in *Arabidopsis*. *Plant Mol Biol* 58: 89-107.
- Leptin M. 1994. Morphogenesis. Control of epithelial cell shape changes. *Curr Biol* 4: 709-712.
- Lin J, Makris A, McMahon C, Bear S, Patriotis C, Prasad V, Brent R, Golemis E, Tschlis P. 1999. The ankyrin repeat-containing adaptor protein Tvl-1 is a novel substrate and regulator of Raf-1. *J Biol Chem* 274: 14706-14715.

- Lloyd JC, Zakhleniuk OV. 2004. Responses of primary and secondary metabolism to sugar accumulation revealed by microarray expression analysis of the *Arabidopsis* mutant, *pho3*. *J Exp Bot* 55: 1221-1230.
- Lu H, Rate D, Song JT, Greenberg J. 2003. ACD6, a novel ankyrin protein, is a regulator and effector of acid salicylic acid signalling in the *Arabidopsis* defense response. *Plant Cell* 15: 2408-2420.
- Lux S, John K, Bennet V. 1990. Analysis of cDNA for human erythrocyte ankyrin indicates a repeated structure with homology to tissue-differentiation and cell-cycle control proteins. *Nature* 344: 36-42.
- Ma L, Sun N, Liu X, Jiao Y, Zhao H, Deng XW. 2005. Organ-specific Expression of *Arabidopsis* Genome during development. *Plant Physiol* 138: 80-91.
- MacKenzie F, Duriez P, Larrivee B, Chang L, Pollet I, Wong F, Yip C, Karsan A. 2004. Notch4-induced inhibition of endothelial sprouting requires the ankyrin repeats and involves signaling through RBP-Jkappa. *Blood* 104: 1760-1768.
- Mansfield S, Briarty L. 1990. Early embryogenesis in *Arabidopsis thaliana* II. The developing embryo. *Can J Bot* 69: 461-476.
- Mantyla E, Lang V, Palva ET. 1995. Role of Abscisic Acid in Drought-Induced Freezing Tolerance, Cold Acclimation, and Accumulation of LT178 and RAB18 Proteins in *Arabidopsis thaliana*. *Plant Physiol* 107: 141-148.
- Massung R, McFadden G, Moyer R. 1992. Nucleotide sequence analysis of a unique near-terminal region of the tumorigenic poxvirus, Shope fibroma virus. *J Gen Vir* 73: 2903-2911.
- Mayer U, Büttner G, Jürgens G. 1993. Apical-basal pattern formation in the *Arabidopsis* embryo: studies on the role of the *GNOM* gene. *Development* 117: 149-162.
- Mayer U, Jurgens G. 1998. Pattern formation in plant embryogenesis: a reassessment. *Semin Cell Dev Biol* 9: 187-193.
- Mayer U, Torres-Ruiz R, Berleth T, Miséra S, Jürgens G. 1991. Mutations affecting body organization in the *Arabidopsis* embryo. *Nature* 353: 402-407.
- Mayfield JA, Fiebig A, Johnstone SE, Preuss D. 2001. Gene families from the *Arabidopsis thaliana* pollen coat proteome. *Science* 292: 2482-2485.
- McElver J, Tzafrir I, Aux G, Rogers R, Ashby C, Smith K, Thomas C, Schetter A, Zhou Q, Cushman MA, Tossberg J, Nickle T, Levin JZ, Law M, Meinke D, Patton D. 2001. Insertional mutagenesis of genes required for seed development in *Arabidopsis thaliana*. *Genetics* 159: 1751-1763.
- Meinke D. 1991. Perspectives on genetic analysis of plant embryogenesis. *Plant Cell* 3: 857-866.
- Meinke D. 1995. Molecular genetics of plant embryogenesis. *Ann Rev Plant Phys Plant Mol Biol* 46: 369-394.
- Meinke DW, Meinke LK, Showalter TC, Schissel AM, Mueller LA, Tzafrir I. 2003. A sequence-based map of *Arabidopsis* genes with mutant phenotypes. *Plant Physiol* 131: 409-418.
- Menges M, de Jager SM, Gruissem W, Murray JA. 2005. Global analysis of the core cell cycle regulators of *Arabidopsis* identifies novel genes, reveals multiple and highly specific profiles of expression and provides a coherent model for plant cell cycle control. *Plant J* 41: 546-566.
- Meyerowitz EM. 1994. Plant developmental biology: green genes for the 21<sup>st</sup> century. *BioEssays* 16: 621-625.
- Michaely P, Bennett V. 1992. The ANK repeat: A ubiquitous motif involved in macromolecular recognition. *Trends Cell Biol* 2: 127-129.
- Michaely P, Tomchick DR, Machius M, Anderson RG. 2002. Crystal structure of a 12 ANK repeat stack from human ankyrinR. *EMBO J* 21: 6387-6396.
- Miner D, Rajkovic A. 2003. Identification of expressed sequence tags preferentially expressed in human placentas by in silico subtraction. *Prenat Diagn* 23: 410-419.

- Miura A, Kato M, Watanabe K, Kawabe A, Kotani H, Kakutani T. 2004. Genomic localization of endogenous mobile CACTA family transposons in natural variants of *Arabidopsis thaliana*. *Mol Genet Genom* 270: 524-532.
- Mosavi L, Minor D, Peng Z. 2002. Consensus-derived structural determinants of the ankyrin repeat motif. *PNAS USA* 99: 16029-16034.
- Mouline K, Véry A, Gaymard F, Boucherez J, Pilot G, Devic M, Bouchez D, Thibaud J, Sentenac H. 2002. Pollen tube development and competitive ability are impaired by disruption of a Shaker K<sup>+</sup> channel in *Arabidopsis*. *Genes Dev* 16: 339-350.
- Niemeyer BA. 2005. Structure-function analysis of TRPV channels. *Naunyn Schmiedeberg Arch Pharm* 371: 285-294.
- Nuccio ML, Thomas TL. 1999. ATS1 and ATS3: two novel embryo-specific genes in *Arabidopsis thaliana*. *Plant Mol Biol* 39: 1153-1163.
- Olsen OA. 2004. Nuclear endosperm development in cereals and *Arabidopsis thaliana*. *Plant Cell* 16 Suppl: S214-227.
- Ostergaard L, Yanofsky MF. 2004. Establishing gene function by mutagenesis in *Arabidopsis thaliana*. *Plant J* 39: 682-696.
- Pan X, Liu H, Clarke J, Jones J, Bevan M, Stein L. 2003. ATIDB: *Arabidopsis thaliana* insertion database. *Nucl Acids Res* 31: 1245-1251.
- Paquette AJ, Benfey PN. 2005. Maturation of the ground tissue of the root is regulated by gibberellin and SCARECROW and requires SHORT-ROOT. *Plant Physiol* 138: 636-640.
- Parcy F, Valon C, Raynal M, Gaubier-Comella P, Delseny M, Giraudat J. 1994. Regulation of gene expression programs during *Arabidopsis* seed development: roles of the ABI3 locus and of endogenous abscisic acid. *Plant Cell* 6: 1567-1582.
- Parinov S, Sundaresan V. 2000. Functional genomics in *Arabidopsis*: large-scale insertional mutagenesis complements the genome sequencing project. *Curr Opin Biotech* 11: 157-161.
- Patthy L. 2003. Modular assembly of genes and the evolution of new functions. *Genetica* 118: 217-231.
- Paxson-Sowders DM, Dodrill CH, Owen HA, Makaroff CA. 2001. DEX1, a novel plant protein, is required for exine pattern formation during pollen development in *Arabidopsis*. *Plant Physiol* 127:1739-1749.
- Peck S, Nühse T, Hess D, Iglesias A, Meins F, Boller T. 2001. Directed proteomics identifies a plant-specific protein rapidly phosphorylated in response to bacterial and fungal elicitors. *Plant Cell* 13: 1467-1475.
- Peleman J, Cottyn B, van Camp W, van Montagu M, Inzé D. 1991. Transient occurrence of extrachromosomal DNA of an *Arabidopsis thaliana* transposon-like element, *Tat1*. *PNAS USA* 88: 3618-3622.
- Peng JB, Brown EM, Hediger MA. 2001. Structural conservation of the genes encoding CaT1, CaT2, and related cation channels. *Genomics* 76: 99-109.
- Pennisi E. 2003. Human genome: reaching their goal early, sequencing labs celebrate. *Science* 300 (5618): 209.
- Pilot G, Pratelli R, Gaymard F, Meyer Y, Sentenac H. 2003. Five-Group distribution of the shaker-like K<sup>+</sup> channel family in high plants. *J Mol Evol* 56: 418-434.
- Ponting CP. 1997. CBS domains in CIC chloride channels implicated in myotonia and nephrolithiasis (kidney stones). *J Mol Med* 75: 160-163.
- Przytycka T, Davis G, Song N, Durand D. 2005. Graph Theoretical Insights into Evolution of Multidomain Proteins. Miyano S et al. (Eds.). RECOMB 2005, pp. 311-325, Springer-Verlag Berlin Heidelberg.
- Rate DN, Cuenca JV, Bowman GR, Guttman DS, Greenberg JT. 1999. The gain-of-function *Arabidopsis* *acd6* mutant reveals novel regulation and function of the salicylic acid signaling pathway in controlling cell death, defenses, and cell growth. *Plant Cell* 11: 1695-1708.



- Redei GP. 1970. *Arabidopsis thaliana* (L.) Heynh. A review of the genetics and biology. *Bibliogr. Genet* 20: 1-151.
- Reiser L, Modrusan Z, Margossian L, Samach A, Ohad N, Haughn G, Fischer R. 1995. The *BELL1* gene encodes a homeodomain protein involved in pattern formation in the *Arabidopsis* ovule primordium. *Cell* 83: 735-742.
- Robinson SJ, Cram DJ, Lewis CT, Parkin IA. 2004. Maximizing the efficacy of SAGE analysis identifies novel transcripts in *Arabidopsis*. *Plant Physiol* 136: 3223-3233.
- Rogers HJ. 2006. Cell death and organ development in plants. *Curr Top Dev Biol* 71: 225-261.
- Rohde K, Bork P. 1993. A fast, sensitive pattern-matching approach for protein sequences. *CABIOS* 9: 183-189.
- Ross JH, Murphy DJ. 1996. Characterization of anther-expressed genes encoding a major class of extracellular oleosin-like proteins in the pollen coat of *Brassicaceae*. *Plant J* 9: 625-637.
- Ross KJ, Franz P, Armstrong SJ, Vizir I, Mulligan B, Franklin FC, Jones GH. 1997. Cytological characterization of four meiotic mutants of *Arabidopsis* isolated from T-DNA-transformed lines. *Chromosome Res* 5: 551-559.
- Rosso MG, Li Y, Strizhov N, Reiss B, Dekker K, Weisshaar B. 2003. An *Arabidopsis thaliana* T-DNA mutagenized population (GABI-Kat) for flanking sequence tag-based reverse genetics. *Plant Mol Biol* 53: 247-259.
- Rubstov A, Lopina O. 2000. Ankyrins. *FEBS Lett* 482: 1-5.
- Rudd S. 2003. Expressed sequence tags: alternative or complement to whole genome sequences? *Trends Plant Sci* 8: 321-329.
- Sabatini S, Heidstra R, Wildwater M, Scheres B. 2003. *SCARECROW* is involved in positioning the stem cell niche in the *Arabidopsis* root meristem. *Genes Dev* 17: 354-358.
- Sambrook J, Fritsch EF, Maniatis T. 1989. *Molecular cloning: a laboratory manual*. Cold Spring Harbor Laboratory Press. Cold Spring Harbor, New York.
- Sanger F, Nicklen S, Coulson AR. 1977. DNA sequencing with chain-terminating inhibitors. *PNAS* 74: 5463-5467.
- Schmid M, Davison TS, Henz SR, Pape UJ, Demar M, Vingron M, Scholkopf B, Weigel D, Lohmann JU. 2005. A gene expression map of *Arabidopsis thaliana* development. *Nat Genet* 37: 501-506.
- Schwacke R, Schneider A, van der Graaff E, Fischer K, Catoni E, Desimone M, Frommer WB, Flugge UI, Kunze R. 2003. ARAMEMNON, a novel database for *Arabidopsis* integral membrane proteins. *Plant Physiol* 131: 16-26.
- Schwartz B, Yeung E, Meinke D. 1994. Disruption of morphogenesis and transformation of the the suspensor in abnormal mutants of *Arabidopsis*. *Development* 120: 3235-3245.
- Scott RJ. 1994. Pollen exine: The sporopollenin enigma and the physics of pattern. In *Molecular and Cellular Aspects of Plant Reproduction*, R.J. Scott and A.D. Stead, eds (Cambridge, UK: Cambridge University Press), pp. 49-81.
- Sedgwick S, Smerdon S. 1999. The ankyrin repeat: a diversity of interactions on a common structural framework. *TIBS* 24: 311-316.
- Sentenac H, Bonneaud, N., Minet M, Lacroute F, Salmo J, Gaymard F, Grignon C. 1992. Cloning and expression in yeast of a plant potassium ion transport system. *Science*. 256: 663-665.
- Sessions A, Burke E, Presting G, aux G, McElver J, Patton D, Dietrich B, Ho P, Bacwaden J, Ko C, Clarke J, Cotton D, Bullis D, Snell J, Miguel T, Hutchinson D, Kimmerly B, Mitzel T, Katagiri F, Glazebrook J, Law M, Goff S. 2002. A high-throughput *Arabidopsis* reverse genetics system. *Plant Cell* 14: 2985-2994.
- Sheen J. 1997. Introduction of plasmid DNA into cells. *ICur. Prot. Mol. Bio.* 1: pp. 1.8.1-1.8.10.

- Sheridan W, Clark J. 1993. Mutational analysis of morphogenesis of the maize embryo. *Plant J* 3: 347-358.
- Sheridan W. 1995. Genes and embryo morphogenesis in angiosperms. *Dev Gen* 16: 291-297.
- Shirley AM, McMichael CM, Chapple C. 2001. The *sng2* mutant of *Arabidopsis* is defective in the gene encoding the serine carboxypeptidase-like protein sinapoylglucose:choline sinapoyltransferase. *Plant J* 28: 83-94.
- Shirley B, Hanley S, Goodman H. 1992. Effects of ionizing radiation on a plant genome: analysis of two *Arabidopsis transparent testa* mutations. *Plant cell* 4: 3333-3347.
- Sintchak MD, Fleming MA, Futer O, Raybuck SA, Chambers SP, Caron PR, Murcko MA, Wilson KP. 1996. Structure and mechanism of inosine monophosphate dehydrogenase in complex with the immunosuppressant mycophenolic acid. *Cell* 85: 921-930.
- Slack JM. 1994. Inducing factors in *Xenopus* early embryos. *Curr Biol* 1: 116-126.
- Souter M, Lindsey K. 2000. Polarity and signalling in plant embryogenesis. *J Exp Bot* 51: 971-983.
- Southern EM. 1975. Detection of specific sequences among DNA fragments separated by gel electrophoresis. *J. Mol. Biol.* 98: 503-517.
- Stiefel V, López E, Roca R, Bastida M, Jahrmann T, Graziano E, Puigdomènech P. 1999. TM20, a gene coding for a new class of transmembrane proteins expressed in the meristematic tissues of maize. *J Biol Chem* 274: 27734-27739.
- Story GM, Peier AM, Reeve AJ, Eid SR, Mosbacher J, Hricik TR, Earley TJ, Hergarden AC, Andersson DA, Hwang SW, McIntyre P, Jegla T, Bevan S, Patapoutian A. 2003. ANKTM1, a TRP-like channel expressed in nociceptive neurons, is activated by cold temperatures. *Cell* 112: 819-829.
- Strasburger E. 1994. *Tratado de Botánica*. 8a. ed. castellana. Ed. Omega. Barcelona.
- Strotmann R, Harteneck C, Nunnenmacher K, Schultz G, Plant T. 2000. OTRPC4, a nonselective cation channel that confers sensitivity to extracellular osmolarity. *Nature Cell Biol* 2: 695-702.
- Sudhof TC. 2001. Alpha-Latrotoxin and its receptors: neurexins and CIRL/latrophilins. *Annu Rev Neurosci* 24: 933-962.
- Takahashi M, Skvarla JJ. 1991. Exine pattern formation by plasma membrane in *Bougainvillea spectabilis* (Nyctaginaceae). *Am J Bot* 78: 1063-1069.
- Tatusov RL, Galperin MY, Natale DA, Koonin EV. 2000. The COG database: a tool for genome-scale analysis of protein functions and evolution. *Nucl Ac Res* 28: 33-36.
- Teichmann SA, Park J, Chothia C. 1998. Structural assignments to the mycoplasma genitalium proteins show extensive gene duplications and domain rearrangements. *PNAS USA* 95: 14658-14663.
- The *Arabidopsis* Genome Initiative, 2000. Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. *Nature* 408: 796-815.
- The *Ceanorhabditis elegans* Sequencing Consortium. 1998. Genome sequence of the *Ceanorhabditis elegans*: a platform for investigating biology. *Science*. 283: 2012-2018.
- Thompson RD, Hueros G, Becker H, Maitz M. 2001. Development and functions of seed transfer cells. *Plant Sci.* 160(5): 775-783.
- Thornycroft D, Sherson SM, Smith SM. 2001. Using gene knockouts to investigate plant metabolism. *J Exp Bot* 52: 1593-1601.
- Toledo-Ortiz G, Huq E, Quail PH. 2003. The *Arabidopsis* Basic/Helix-Loop-Helix Transcription Factor Family. *Plant Cell* 15: 1749-1770.
- Tordai H, Nagy A, Farkas K, Banyai L, Patthy L. 2005. Modules, multidomain proteins and organismic complexity. *FEBS J* 272: 5064-5078.

- Torres-Ruiz R, Lohner A, Jürgens G. 1996. The GURKE gene is required for normal organization of the apical region in the Arabidopsis embryo. *Plant J* 10: 1005-1016.
- Tsay Y, Frank M, Page T, Dean C, Crawford N. 1993. Identification of a mobile endogenous transposon in *Arabidopsis thaliana*. *Science* 260: 342-344.
- Tzafrir I, Pena-Muralla R, Dickerman A, Berg M, Rogers R, Hutchens S, Sweeney TC, McElver J, Aux G, Patton D, Meinke D. 2004. Identification of genes required for embryo development in Arabidopsis. *Plant Physiol* 135: 1206-1220.
- Ueda K, Tanaka I. 1995. The appearance of male gamete-specific histones gH2B and gH3 during pollen development in *Lilium longiflorum*. *Dev Biol* 169: 210-217.
- Unoki M, Nakamura Y. 2001. Growth-suppressive effects of BPOZ and EGR2, two genes involved in the PTEN signaling pathway. *Oncogene* 20: 4457-4465.
- Urbánek P, Pačes J, Pačes V. 2005. An approach towards experimental cDNA sequence determination of predicted genes: an example from Arabidopsis U3-55k homologues. *Gene* 358: 67-72.
- Van Lijsebettens M, Van Montagu, M. 2005. Historical perspectives on plant development biology. *Int J Dev Biol* 49: 453-465.
- Vasmatazis G, Essand M, Brinkmann U, Lee B, Pastan I. 1998. Discovery of three genes specifically expressed in human prostate by expressed sequence tag database analysis. *PNAS USA*: 300-304.
- Venter J, *et al.*, 2001. The sequence of the human genome. *Science* 291 (5507): 1304-1351.
- Vernon D, Hannon M, Le M, Forsthoefel N. 2001. An expanded role for the *TWIN1* gene in embryogenesis: defects in cotyledon pattern and morphology in the *twin1* mutant of Arabidopsis (Brassicaceae). *Am J Bot* 88: 570-582.
- Vernon D, Meinke D. 1994. Embryonic transformation of the suspensor in *twin*, a polyembryonic mutant of Arabidopsis. *Dev Biol* 165: 566-573.
- Vicient CM, Delseny M. 1999. Isolation of total RNA from *Arabidopsis thaliana* seeds. *Anal Biochem* 268: 412-413.
- Vicient CM, Hull G, Guilleminot J, Devic M, Delseny M. 2000. Differential expression of the Arabidopsis genes coding for Em-like proteins. *J Exp Bot* 51: 1211-1220.
- Vogel C, Bashton M, Kerrison ND, Chothia C, Teichmann SA. 2004. Structure, function and evolution of multidomain proteins. *Curr Opin Struct Biol* 14: 208-216.
- Ward J. 2001. Identification of novel families of membrane proteins from the model plant *Arabidopsis thaliana*. *Bioinformatics* 17 (6): 560-563.
- Welle S, Bhatt K, Thornton CA. 1999. Inventory of high-abundance mRNAs in skeletal muscle of normal men. *Genome Res* 9: 506-513.
- West M, Harada J. 1993. Embryogenesis in higher plants: an overview. *Plant Cell* 5: 1361-1369.
- West MAL, Yee KM, Danao J, Zimmerman JL, Fischer RL, Goldberg RB, Harada JJ. 1994. LEAFY COTYLEDON1 Is an Essential Regulator of Late Embryogenesis and Cotyledon Identity in Arabidopsis. *Plant Cell* 6: 1731-1745.
- White JA, Todd J, Newman T, Focks N, Girke T, Martínez de Llárduya O, Jaworski JG, Ohlrogge JB, Benning C. 2000. A New Set of Arabidopsis Expressed Sequence Tags from Developing Seeds. The Metabolic Pathway from Carbohydrates to Seed Oil. *Plant Physiol* 124: 1582-1594.
- Wigge P, Weigel D. 2001. Arabidopsis genome: life without notch. *Curr Biol* 11: R112-R114.
- Wilhelm KS, Thomashow MF. 1993. Arabidopsis thaliana cor15b, an apparent homologue of cor15a, is strongly responsive to cold and ABA, but not drought. *Plant Mol Biol* 23: 1073-1077.

- Willemsen V, Scheres B. 2004. Mechanisms of pattern formation in plant embryogenesis. *Annu Rev Genet* 38: 587-614.
- Wilson ZA, Yang C. 2004. Plant gametogenesis: conservation and contrasts in development. *Reproduction* 128: 483-492.
- Xu J, Zhang HY, Xie CH, Xue HW, Dijkhuis P, Liu CM. 2005. EMBRYONIC FACTOR 1 encodes an AMP deaminase and is essential for the zygote to embryo transition in Arabidopsis. *Plant J*. 42(5): 743-756.
- Yadegari R, Paiva G, Laux T, Koltunow A, Apuya N, Zimmerman J, Fischer R, Harada J, Goldberg R. 1994. Cell differentiation and morphogenesis are uncoupled in Arabidopsis raspberry embryos. *Plant Cell* 6 (12): 1713-1729.
- Yamada K, Lim J, Dale JM, Chen H, Shinn P, Palm CJ, Southwick AM, Wu HC, Kim C, Nguyen M, Pham P, Cheuk R, Karlin-Newmann G, Liu SX, Lam B, Sakano H, Wu T, Yu G, Miranda M, Quach HL, Tripp M, Chang CH, Lee JM, Toriumi M, Chan MM, Tang CC, Onodera CS, Deng JM, Akiyama K, Ansari Y, Arakawa T, Banh J, Banno F, Bowser L, Brooks S, Carninci P, Chao Q, Choy N, Enju A, Goldsmith AD, Gurjal M, Hansen NF, Hayashizaki Y, Johnson-Hopson C, Hsuan VW, Iida K, Karnes M, Khan S, Koesema E, Ishida J, Jiang PX, Jones T, Kawai J, Kamiya A, Meyers C, Nakajima M, Narusaka M, Seki M, Sakurai T, Satou M, Tamse R, Vaysberg M, Wallender EK, Wong C, Yamamura Y, Yuan S, Shinozaki K, Davis RW, Theologis A, Ecker JR. 2003. Empirical analysis of transcriptional activity in the Arabidopsis genome. *Science* 302: 842-846.
- Yan J, Wang J, Zhang H. 2002. An ankyrin repeat-containing protein plays a role in both disease resistance and antioxidation metabolism. *Plant J* 29: 193-202.
- Yang Y, Nanduri S, Sen S, Qin J. 1998. The structural basis of Ankyrin-like repeat function as revealed by the solution structure of myotrophin. *Structure* 6: 619-626.
- Yeung E, Meinke D. 1993. Embryogenesis in angiosperms: development of the suspensor. *Plant Cell* 5: 1371-1381.
- Yu J, Hu S, Wang J, Wong GK, Li S, Liu B, Deng Y, Dai L, Zhou Y, Zhang X, Cao M, Liu J, Sun J, Tang J, Chen Y, Huang X, Lin W, Ye C, Tong W, Cong L, Geng J, Han Y, Li L, Li W, Hu G, Huang X, Li W, Li J, Liu Z, Li L, Liu J, Qi Q, Liu J, Li L, Li T, Wang X, Lu H, Wu T, Zhu M, Ni P, Han H, Dong W, Ren X, Feng X, Cui P, Li X, Wang H, Xu X, Zhai W, Xu Z, Zhang J, He S, Zhang J, Xu J, Zhang K, Zheng X, Dong J, Zeng W, Tao L, Ye J, Tan J, Ren X, Chen X, He J, Liu D, Tian W, Tian C, Xia H, Bao Q, Li G, Gao H, Cao T, Wang J, Zhao W, Li P, Chen W, Wang X, Zhang Y, Hu J, Wang J, Liu S, Yang J, Zhang G, Xiong Y, Li Z, Mao L, Zhou C, Zhu Z, Chen R, Hao B, Zheng W, Chen S, Guo W, Li G, Liu S, Tao M, Wang J, Zhu L, Yuan L, Yang H. 2002. A draft sequence of the rice genome (*Oryza sativa* L. ssp. *indica*). *Science* 296: 79-92.
- Zhang B, Peng Z. 2000. A minimum folding unit in the ankyrin repeat protein p16 (INK4). *J Mol Biol* 299: 1121-1132.
- Zhang H, Scheirer D, Fowle W, Goodman H. 1992. Expression of antisense or sense RNA of an ankyrin repeat-containing gene blocks chloroplast differentiation in Arabidopsis. *Plant Cell* 4: 1575-1588.
- Zhang J, Somerville C. 1997. Suspensor-derived polyembryony caused by altered expression of valyl-tRNA synthetase in the *tnw2* mutant of Arabidopsis. *PNAS USA* 94: 7349-7355.
- Zimmermann P, Hennig L, Grissem W. 2005. Gene-expression analysis and network discovery using Genevestigator. *Trends Plant Sci* 10: 407-409.
- Zimmermann P, Hirsch-Hoffmann M, Hennig L, Grissem W. 2004. GENEVESTIGATOR. Arabidopsis Microarray Database and Analysis Toolbox. *Plant Physiol* 136: 2621-2632.
- Zimmermann S, Talke I, Ehrhardt T, Nast G, Müller-Röber B. 1998. Characterization of SKT1, an inwardly rectifying potassium channel from potato, by heterologous expression in insect cells. *Plant Physiol* 116: 879-890.

## **APÉNDICE I**

---

### **Computational and experimental analysis identifies Arabidopsis genes specifically expressed during early seed development**

Cristian Becerra, Pere Puigdomènech y Carlos M. Vicent

BMC Genomics (2006) 7: 38

Methodology article

Open Access

## Computational and experimental analysis identifies Arabidopsis genes specifically expressed during early seed development

Cristian Becerra, Pere Puigdomenech and Carlos M Vicient\*

Address: Laboratori de Genetica Molecular i Vegetal, CSIC-IRTA, Jordi Girona 18–36, 08034, Barcelona, Spain

Email: Cristian Becerra - cbbgmp@cid.csic.es; Pere Puigdomenech - pprgmp@cid.csic.es; Carlos M Vicient\* - cvsmp@cid.csic.es

\* Corresponding author

Published: 28 February 2006

Received: 10 October 2005

BMC Genomics 2006, 7:38 doi:10.1186/1471-2164-7-38

Accepted: 28 February 2006

This article is available from: <http://www.biomedcentral.com/1471-2164/7/38>

© 2006 Becerra et al; licensee BioMed Central Ltd.

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/2.0>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

### Abstract

**Background:** Plant seeds are complex organs in which maternal tissues, embryo and endosperm, follow distinct but coordinated developmental programs. Some morphogenetic and metabolic processes are exclusively associated with seed development. The goal of this study was to explore the feasibility of incorporating the available online bioinformatics databases to discover Arabidopsis genes specifically expressed in certain organs, in our case immature seeds.

**Results:** A total of 11,032 EST sequences obtained from isolated immature seeds were used as the initial dataset (178 of them newly described here). A pilot study was performed using EST virtual subtraction followed by microarray data analysis, using the Geneinvestigator tool. These techniques led to the identification of 49 immature seed-specific genes. The findings were validated by RT-PCR analysis and *in situ* hybridization.

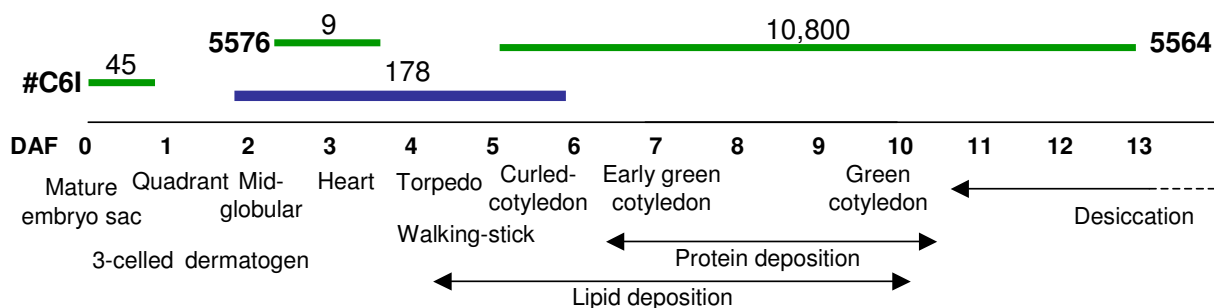
**Conclusion:** We conclude that the combined *in silico* data analysis is an effective data mining strategy for the identification of tissue-specific gene expression.

### Background

Seeds are complex genetic entities with a diploid maternal genotype, derived from the ovary wall, a diploid embryo, with equal genetic contributions from the pollen donor and pollen recipient, and a triploid endosperm, in which the maternal genetic contribution is twice that of the paternal parent. Endosperm development is a process with many unique features determining the coordinated development and disappearance of a highly specialized organ [1]. During embryogenesis, the egg cell divides and develops into an embryo, passing through different developmental phases: globular, heart, torpedo, cotyledon, curled-cotyledon and maturation [2]. Key steps in early embryo development are the acquisition of a polar structure with a shoot-root axis, the formation of the apical and root meristems, and the differentiation of the cotyle-

don primordia. After this last stage, the size of the embryo increases and deposition of storage macromolecules begins. Finally, during maturation, the embryo desiccates. During this process, the seed coat develops from the two integuments that surround the embryo. Several of the processes described above are not present in any other plant tissues, so the genetic program for seed development is likely to involve the concerted activity of many seed-specific genes.

Determination of the genes involved in seed development, and their functions, is one of the major goals in plant developmental biology. Mutational approaches have been extensively used to analyse seed development in Arabidopsis [3-5]. Several mutants have been isolated giving loss-of- or altered-seed development allowing the



**Figure 1**  
**Overview of EST libraries from isolated immature *Arabidopsis* seeds.** At the top, a representation of the available EST collections extracted from immature seeds. Lines in colour represent the period of development covered by the library. The library code according to the TIGR *Arabidopsis* Gene Index ([http://www.tigr.org/tigr-scripts/tgi/T\\_index.cgi?species=arab](http://www.tigr.org/tigr-scripts/tgi/T_index.cgi?species=arab) [29]) is indicated next to the line. The number of ESTs available from the corresponding library is indicated above the line. Green lines correspond to previously existing EST collections, and the blue line corresponds to the new library described here. At the bottom, the stages of embryo and seed development, related to days after flowering (DAF), is shown [49]. The main processes associated with seed development are indicated.

identification of several genes [6,7]. However, insertional mutagenesis has some deficiencies. For example, probably due to gene redundancy, many of the insertions in genes do not produce any detectable phenotype, and genes whose disruption produces alterations in seed development are not necessarily genes with seed specific expression [6]. In consequence, although mutational approaches have been, and still are, basic for understanding the processes involved in seed development, they are not enough to build a complete picture of the process.

Expression profiling and definition of genes specifically or preferentially expressed in certain tissues complement the genetic and molecular approaches. The generation of EST collections and the oligonucleotide-based microarrays can produce reliable, high-quality data [8,9]. The deposition of the results of RNA profiling experiments in public databases provides a valuable tool for *in silico* analysis of organ specific gene expression. There have been several reports of EST-based computer analysis of human tissue transcriptomes [10-15], and computer analyses have been performed in differential human EST database searches [16].

EST abundance in plants is not as high as for humans, but for some species the total number of ESTs in publicly available databases exceeds the total number of genes by more than one order of magnitude. For example, the NCBI dbEST database release 111105 (November 11, 2005) [17] included 656,945 from *Zea mays* (maize), 600,039 sequences from *Triticum aestivum* (wheat),

420,789 from *Arabidopsis thaliana* (thale cress) and 406,790 from *Oryza sativa* (rice), compared with the 7,057,754 for humans. Despite this, there are few examples of *in silico* expression studies in plants [18,19].

From the complete sequencing of certain plant genomes, it is possible to monitor gene expression on a genome-scale using high-density oligonucleotide arrays [20]. Thousands of *Arabidopsis* arrays, containing probes for more than twenty thousand genes, have been processed, and systematic analyses of gene expression in different organs, developmental conditions and stress responses, have been performed [9,21-23]. The results of many of these are publicly available through web browser interfaces such as the Genevestigator tool [24-26]. In view of this, at least for *Arabidopsis*, data analysis rather than data collection is the first challenge for biologists in determining patterns of gene expression.

The focus of this work was the identification of genes whose expression is specific in immature seeds. Firstly, we sequenced cDNA clones from isolated immature seeds. Secondly, we used *in silico* subtraction in a combination of EST selection and microarray data analysis in order to select genes with the desired pattern of expression. Finally, 49 genes specifically expressed during seed development were selected. Our study demonstrates the reliability of *in silico* subtraction methods in *Arabidopsis* and provides a basis for targeted reverse-genetic approaches aimed at identifying key genes involved in reproductive development in plants.

**Table 1: Genes selected by *in silico* subtraction**

Gene AGI code	Imm. seed ESTs	Indifferent ESTs	Definition	Functional category	Pattern of expression <sup>1</sup>	Mutants	Tandem arrays	Segmental duplication
At1g03790	1	5	Zinc finger (CCCH-type) family protein	Regulation of gene expression	llc	-	1	1
At1g03890	41	22	Cruciferin 12S seed storage protein	Nutrient reservoir	llb	-	2	1
At1g14950	2	15	Major latex protein type I	Secondary metabolism	llc	-	4	1
At1g48130	2	12	Peroxiredoxin	Response to abiotic stress	llb	-	1	1
At1g48660	1	0	Auxin-responsive GH3 family protein	Development	llc	-	3	1
At1g62060	32	25	Unknown	Unknown	lla	-	2	1
At1g65090	2	6	Unknown	Unknown	llb	-	1	1
At1g67100	3	3	Seed specific protein Bn15D17A	Unknown	llb	-	1	1
At1g73190	8	16	Tonoplast intrinsic protein 3.1	Protein processing	llb	-	1	2
At1g80090	3	2	Unknown	Unknown	llb	-	1	1
At2g28420	1	4	Lactoylglutathione lyase family protein	Carbohydrate metabolism	llc	-	1	1
At2g33520	1	1	Unknown	Unknown	llc	-	1	1
At2g34700	2	4	Proline-rich glycoprotein	Development	l	-	1	1
At3g01570	15	52	Oleosin	Nutrient reservoir	llb	-	1	1
At3g04170	1	0	Germin-like protein subfamily I	Unknown	l	-	5	1
At3g04190	1	0	Germin-like protein subfamily I	Unknown	l	-	5	1
At3g12960	1	0	Similar to seed maturation protein PM28	Unknown	llc	-	1	1
At3g24650	6	3	ABI3 protein	Regulation of gene expression	llb	Abi3 <sup>2</sup>	1	1
At3g27660	7	0	Oleosin	Nutrient reservoir	llb	-	1	1
At3g48580	1	1	Xyloglucan:xyloglucosyl transferase	Carbohydrate metabolism	llc	-	1	1
At3g54940	4	18	Cysteine proteinase	Protein processing	llb	-	1	1
At3g60730	2	2	Pectinesterase-like protein	Development	llc	-	1	1
At3g61040	1	0	Cytochrome P450 monooxygenase-like	Respiration and energy	llc	-	1	1
At3g62730	55	17	Desiccation-related protein	Response to abiotic stress	llb	-	1	1
At3g63040	1	0	Unknown	Unknown	llb	-	1	1
At4g25140	1	5	Glycine-rich protein/ oleosin	Nutrient reservoir	llb	-	1	1
At4g27150	68	33	2S seed storage protein 2 precursor	Nutrient reservoir	llb	-	4	1
At4g28520	92	4	12S cruciferin seed storage protein (CRU3)	Nutrient reservoir	llb	-	1	1
At4g36700	48	16	Globulin-like protein	Nutrient reservoir	lla	-	1	1
At4g37050	2	5	Patatin-like	Nutrient reservoir	lla	-	3	1
At5g01670	1	1	Aldose reductase-like protein	Carbohydrate metabolism	llc	-	1	1
At5g03860	1	18	Malate synthase	Carbohydrate metabolism	llc	-	1	1
At5g04010	1	0	Unknown	Unknown	llc	-	1	1
At5g07190	10	15	Embryo-specific protein 3 (ATS3)	Unknown	llb	-	1	1
At5g09640	10	4	Serine carboxypeptidase-like	Protein processing	l	Sng2 <sup>3</sup>	1	1
At5g22470	8	1	Poly (ADP-ribose) polymerase family protein	Protein processing	llc	-	1	1
At5g40420	39	68	Oleosin	Nutrient reservoir	llb	-	1	1
At5g44310	5	1	Late embryogenesis abundant protein-like	Response to abiotic stress	llc	-	1	1
At5g45690	4	6	Unknown	Unknown	llc	-	1	1
At5g45830	1	1	Unknown	Unknown	llc	-	1	1
At5g48100	30	19	Laccase	Response to abiotic stress	lla	-	1	1



**Table 1: Genes selected by *in silico* subtraction (Continued)**

At5g49190	9	0	Sucrose synthase (SUS2)	Carbohydrate metabolism	I	-	1	1
At5g50700	9	41	11-beta-hydroxysteroid dehydrogenase-like	Response to abiotic stress	IIb	-	2	1
At5g54740	7	37	2S storage protein-like	Nutrient reservoir	IIb	-	1	1
At5g55240	6	3	Embryo-specific protein I	Unknown	IIb	-	1	1
At5g57260	1	0	Cytochrome P450	Respiration and energy	IIb	-	1	2
At5g59170	11	6	Cell wall protein precursor, extensin	Development	IIb	-	1	1
At5g62490	2	5	AtHVA22b	Response to abiotic stress	IIc	-	1	1
At5g62800	1	0	Seven in absentia (SINA) family protein	Protein processing	IIb	-	1	1

(1) Information in Figure 4.

(2) Mutant is abscisic acid-insensitive and lacks seed dormancy.

(3) Mutant accumulates sinapoylglucose instead of sinapoylcholine.

## Results and discussion

### Sequencing *Arabidopsis* young seed ESTs

ESTs from isolated *Arabidopsis* immature seeds are not very abundant in EST databases (Figure 1). Among the 420,789 *Arabidopsis* ESTs deposited (release 111105) [17], 10,854 correspond to isolated immature seeds, 10,800 correspond to seeds in mid-development stages [27] and only 54 were obtained from early stages of seed development. We constructed a cDNA library from developing *Arabidopsis* seeds isolated at a stage from mid-globular to curled-cotyledon (2 to 6 days after pollination) and obtained 178 single pass 5' end sequences (>140 bp). The average sequence length was 579 bp. Newly sequenced ESTs were assembled in contigs and gene identities were assigned querying against the *Arabidopsis* genome database at TAIR [28] using the BLAST algorithm. They corresponded to 95 individual genes: 93 nuclear and two from chloroplasts. Functional categories were determined based on GO data in the TAIR database [28]. 21% of the genes are linked to translation, 6% to carbohydrate metabolism and 5% to development. The function of 31% of the genes remained unknown. For two of the genes (At1g60987 and At2g02490) no ESTs have been previously sequenced.

### Identification of genes specifically expressed in seeds during early development

A two step *in silico* subtraction procedure was used to select genes specifically transcribed in immature seeds. The first selection step was based on EST abundance and the second step on microarray data analysis.

The objective of the first step was to identify genes having ESTs only from immature seeds and not from other organs. We divided the *Arabidopsis* EST libraries deposited in the TIGR *Arabidopsis* Gene Index [29] into three categories, according to the organs they were made from (Additional file 1):

a) Immature seed: this includes 10,854 ESTs from four cDNA libraries (Figure 1).

b) Other tissues: this includes 50,992 ESTs from 78 cDNA libraries obtained from vegetative tissues, non-pollinated flowers and dry seeds.

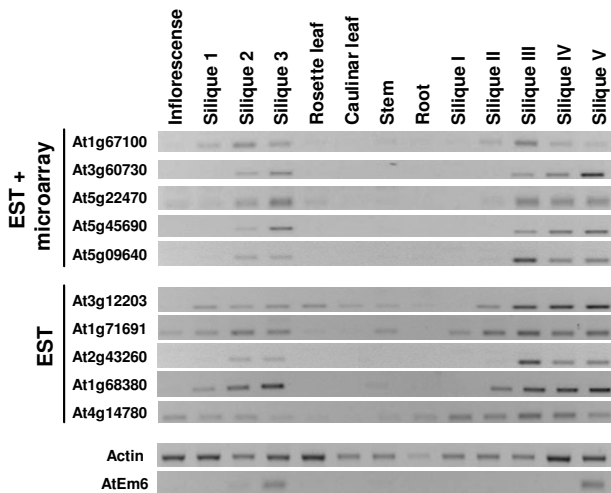
c) Non-informative: this includes libraries obtained from mixed organs and whole plants, including libraries from siliques.

Subtraction was done based on the EST contigs and gene assignments in TIGR *Arabidopsis* Gene Index [29]. We selected genes having corresponding EST sequences in category a (immature seeds) and not in category b (other tissues). 640 genes passed our first subtraction criteria (Additional file 2). Two correspond to chloroplast genes, three to mitochondrial genes and 26 had homology to parts of the *Arabidopsis* genome in which no genes have been reported.

The second selection step was based on the *Arabidopsis* Affymetrix GeneChip® average data available on the Genevestigator analysis tool site [24-26]. We used the meta-analyzer program, which performs a heat map of normalized signal intensity values, corresponding to the different organs of the plant, for each gene. Values range from 0 to 100, 100 being the highest level of expression. We selected the genes using the following criteria:

(i) The expression in seeds should be higher than 80.

(ii) The expression in other organs should be lower than 5, except for siliques, carpels and inflorescences, as these three organs could contain immature seeds at the very early stages after pollination. Detected level 5 is probably low, but was chosen in order to avoid possible errors in



**Figure 2**  
**RT-PCR analysis of the expression profiles of ten genes isolated by in silico screening.** "EST + microarray" indicates genes isolated by the combination of EST selection and microarray data analyses. "EST" indicates genes isolated only by EST selection. Siliques I to 3 correspond to whole siliques at different stages of development (I, young green; 2, green fully developed; 3, desiccating siliques). Siliques I to V correspond to siliques at different stages of development (I, 0–4 daf; II, 4–8 daf; III, 8–12 daf; IV, 12–16 daf; V, 17–21 daf). In each case, the size of the bands was as expected.

the normalisation algorithm in the meta-analyzer program.

(iii) The expression level in seeds should be higher or equal to the expression in siliques, carpels or inflorescences.

49 of the 634 selected genes were not considered in the second analysis because they are not included in the Arabidopsis Affymetrix 22K GeneChip®. Of the remaining 585 genes, 49 (8%) fulfilled the selection criteria and may represent genes specifically expressed in immature seeds (Table 1). From the non-selected genes, 51% did not fit the selection condition (i), 96% the selection condition (ii) and 35% the selection condition (iii). Surprisingly, 21% of the genes showed higher values in siliques than in seeds. The different conditions in which tissues were collected for cDNA synthesis and microarray hybridizations could explain these results.

The advantage of the selection method is demonstrated by the presence of several genes already characterized as specifically expressed in seeds, such as: *abi3* [30]; At1g48130, encoding a peroxiredoxin (PER1) whose expression is

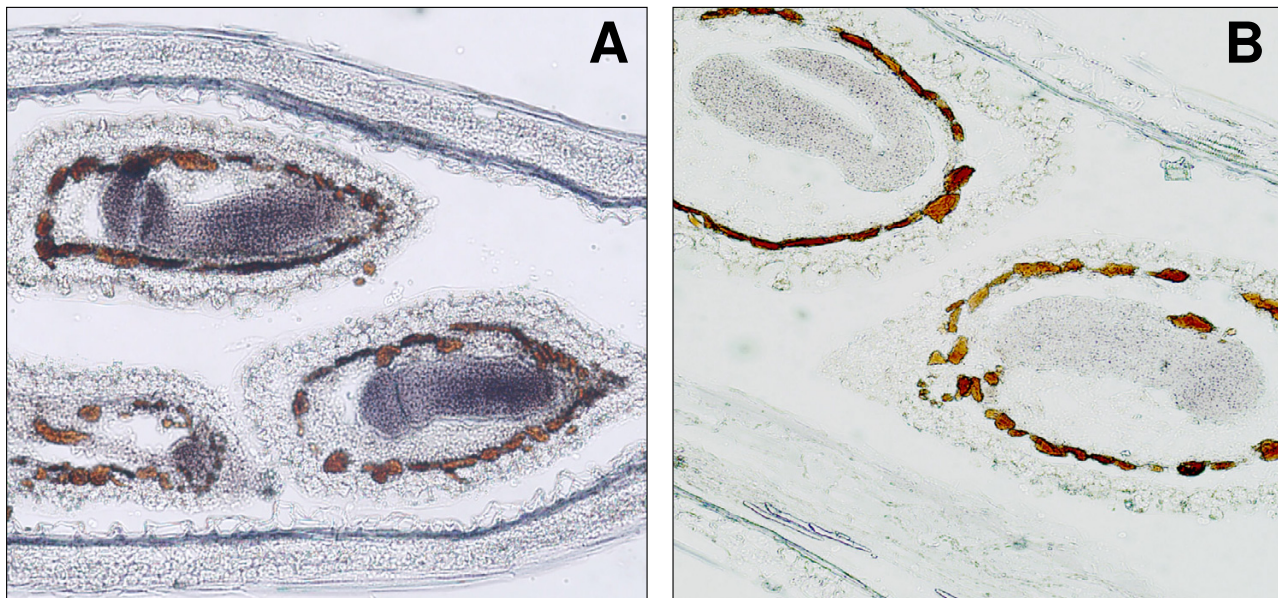
restricted to seeds [31]; At1g67100, which is homologous to the *Brassica Bn15D17A* gene, highly and specifically expressed in embryos and seed coat at the early stages of seed development [32]; and At5g07190 and At5g55240, which encode embryo-specific proteins isolated in the course of a differential display experiment [33].

We also tested the direct application of the microarray subtraction without EST selection. We chose the first 1,500 genes from chromosome 1 (according to the AGI code) included in the Arabidopsis Affymetrix 22K GeneChip® (from At1g01010 to At1g18340). 28 of the 1,500 genes (1.9%) fell within the microarray-based selection criteria. If there is the same proportion in the whole genome, about 550 genes would be selected. These results indicate that Genevestigator may be a useful tool to investigate organ specific gene expression in Arabidopsis. However, data obtained from Genevestigator is based on the normalised average signal intensity values obtained from several array experiments [24-26]. The normalisation algorithms used to generate Genevestigator values could introduce false positives and negatives, particularly for genes with low levels of expression. In consequence, combining Genevestigator results with EST abundance data gives a more reliable dataset of genes specifically expressed in a certain organ, seeds in our case.

#### **Experimental validation of the patterns of expression of the selected genes**

We used RT-PCR to check our selection procedure (Figure 2). Ten genes were selected, five of which were only used in the EST based selection and not the microarray, and the other five genes passed both selection steps. Two genes were used as additional controls: actin, which is expressed in all tissues, and *AtEm6*, which is specifically expressed during late embryogenesis [34]. All 10 genes analyzed showed higher expression levels in siliques, but silique specificity is, in general, higher in the genes selected by EST and microarray than in the genes selected only by EST subtraction. Two of the genes in the EST and microarray group, At1g67100 and At5g22470, gave low levels of amplification in rosette leaves and At1g67100 also in stem. This difference between Genevestigator and experimental data could be a consequence of different levels of detection in RT-PCR and microarray experiments or different experimental conditions. They do not indicate strong bias in the results. EST and microarray based selection produces a specific, expression-based, list of genes.

Seed specific expression was further demonstrated by *in situ* hybridization for the At5g22470 gene encoding a Poly (ADP-ribose) polymerase family protein (PARP) (Figure 3). The At5g22470 transcripts were detected specifically in the embryo and not in the endosperm, pericarp, valves or septum. The profile of the expression of the At5g22470



**Figure 3**

***In-situ* hybridization analysis of a seed-specifically expressed gene.** Seed-specific transcript labelling of embryos at the late torpedo stage as shown by *in situ* hybridization of transverse sections of *Arabidopsis* siliques probed with digoxigenin-labelled *At5g22470* mRNA, viewed under bright-field optics.

gene is consistent with the predicted seed specific transcription.

The RT-PCR experiments and the presence of genes known to be specifically expressed in seed demonstrate that the selection procedure identifies genes specifically, or at least, predominantly, expressed in developing seeds. The relatively low number of genes selected is probably a consequence of the small number of initial ESTs corresponding to immature seeds (11,032 sequences). This is especially true in the case of genes only expressed during very early stages of seed development, for which only 232 ESTs are available. A recent report showed that only 16,115 of *Arabidopsis* genes are represented in the EST databases [35]. An additional problem is that not all the genes are represented in the Affymetrix 22K GeneChip®. We estimate that, if all genes were present in EST and microarray databases about a hundred would have been selected by our *in silico* method. It has been proposed that the developmental processes occurring during embryogenesis are active during the vegetative development of the plant, therefore some genes may also be expressed in other growing organs of the plant, and so not seed specific.

#### **Functional classification of the selected genes**

The 49 selected seed-specific genes were grouped into different functional categories (Table 2) according to their predicted gene products, based on the Gene Ontology (GO) Consortium through the *Arabidopsis* consortium information [28]. The data were compared with the functional categories assigned for all *Arabidopsis* genes [36].

14 of the selected genes correspond to genes of unknown function (28.6%). This is lower but not significantly different (Fisher's exact test,  $\alpha = 0.05$ ) to the percentage obtained for the total genome (38.4%). Particularly interesting is *At1g62060*, whose function is unknown but is represented in databases by a total of 57 EST sequences (32 from immature seed libraries). Two of the genes encode germin-like proteins (*At3g04170* and *At3g04190*), and four have been listed as seed or embryo specific genes of unknown function (*At1g67100*, *At3g12960*, *At5g07190* and *At5g55240*).

Genes in the "nutrient reservoir" category represent 20.4% of the selection and include ten genes, four encoding oleosins, three globulins, two cruciferins and one a patatin-like protein. Accumulation of seed storage proteins is a

highly seed specific process [37], so it is not surprising that the proportion of these genes in the selected group is significantly higher than that obtained for the whole genome (0.2%).

The third category is "response to abiotic stress", which includes six genes (12.2%), and is significantly more abundant than in the whole genome (3.1%). This is an indication of the importance of genes providing stress-tolerance in correct seed development. Three of the genes encode oxidative stress-related enzymes, the function of two genes is related to desiccation (At3g62730 and At5g44310), and one is an ABA and stress inducible gene (At5g62490).

Five genes involved in carbohydrate metabolism were selected (10.2%). This percentage is significantly higher than that observed for the whole genome (2.4%). This category includes a gene encoding a xyloglucan:xyloglucosyl transferase (At3g48580), an enzyme (E.C.2.4.1.207) involved in the biosynthesis of the cell wall. It also includes a gene encoding a sucrose synthase (At5g49190). Sucrose represents a signal for differentiation during embryo development and up-regulates storage-associated gene expression [38].

Five genes involved in protein modification, localization or degradation were selected (10.2%), two of them being proteases (At3g54940 and At5g09640). No genes involved in translation were selected, even though these represent 2.7% of the genes in the whole genome, nor any involved in transport and subcellular trafficking, even

though these represent 8.7% of the genes in the whole genome.

Four genes involved in different aspects of development (8%) were selected. Two of them are involved in cell wall synthesis or modification (At5g59170, encoding a cell wall protein precursor, extensin; and At3g60730, encoding a pectinesterase-like protein). This is an indication of the high rate of synthesis of new cell wall during seed development, and could also be an indication of the importance of specific cell wall components in co-ordinating gene expression programmes during embryo development [39], an effect observed in immature maize embryos [40]. The number of selected genes involved in development is not significantly higher than in the whole genome (60%). This is not surprising as the whole genome contains several genes involved, for example, in flower or root development. A third gene encodes an auxin-responsive GH3 family protein (At1g48660). Auxins are important signalling molecules involved in shoot/root axis establishment, among other processes [41].

Two genes involved in the regulation of gene expression (40%) were selected: *abi3* and a gene encoding a CCCH-type zinc finger protein (At1g03790). Although not significantly, this number is lower than that observed for the whole genome (7.4%). The reduced number of transcription factor genes selected is surprising, but recent data from global analysis of gene expression indicate that the number of transcription factor genes specifically expressed during seed development is relatively low compared with other organs [8,42]. The expression of several

**Table 2: Functional categories of the seed specific genes**

Functional category	Whole genome (%)	Subtracted genes (%) (p-value) <sup>1</sup>
Amino acid metabolism	0.1	0.0 <sup>1.00</sup>
Carbohydrate metabolism	2.4	10.2 <sup>0.01*</sup>
Cell division cycle	2.3	0.0 <sup>0.63</sup>
Defense	0.9	0.0 <sup>1.00</sup>
Development	6.0	8.2 <sup>0.54</sup>
Lipid metabolism	0.9	0.0 <sup>1.00</sup>
Metabolism	6.4	0.0 <sup>0.07</sup>
Nucleic acid metabolism	3.1	0.0 <sup>0.41</sup>
Nutrient reservoir	0.2	20.4 <sup>0.00*</sup>
Photosynthesis	0.3	0.0 <sup>1.00</sup>
Protein processing	9.4	10.2 <sup>0.81</sup>
Regulation of gene expression	7.4	4.1 <sup>0.58</sup>
Respiration and energy	4.0	4.1 <sup>1.00</sup>
Response to abiotic stress	3.1	12.2 <sup>0.00*</sup>
Secondary metabolism	0.7	2.0 <sup>0.28</sup>
Transport and subcellular trafficking	8.7	0.0 <sup>0.02*</sup>
Transcription and splicing	6.1	0.0 <sup>0.07</sup>
Translation	2.7	0.0 <sup>0.64</sup>
Unknown	38.4	28.6 <sup>0.17</sup>

1. p-value for the same or a stronger association of Fisher's exact test compared with total genome

\*. p-value < 0.05.

MADS-box genes have been analyzed in different Arabidopsis tissues and it was found that, although many of these genes are expressed in embryonic tissue culture, few of them are exclusively expressed in this tissue [42]. Similarly, the number of specifically expressed transcription factor genes in developing siliques is relatively low compared to other tissues [8]. An additional explanation could be that, as this category of genes has relatively low levels of expression, they may be under-represented in EST collections used for selection.

Finally, two genes involved in respiration and energy (4.1%) and one in secondary metabolism (2.0%) (At1g14950 encoding a major latex protein type 1) were selected. Interestingly, two of the most highly represented categories in the genome are not represented in our selection: metabolism (6.4%) and transcription and splicing (6.1%). Nor were any genes detected for cell division, metabolism of amino acids, nucleic acid or lipids, defense or photosynthesis. As these genes are involved in general cell processes, they are expressed in several tissues and organs and they are unlikely to be selected in a seed-specific subtraction.

#### **Gene redundancy and mutant phenotypes**

Mutational approaches have been extensively used in Arabidopsis to identify gene functions [3]. Mutation in about 800 genes produced loss of function phenotypes in Arabidopsis [6]. Of these, about 250 produce an altered embryo. Based on the information available in the Arabidopsis information resource (TAIR) [28] and Seedgenes [7], two of the 49 genes have a mutant phenotype (4%) (Table 1), and in only one of them the mutation produces alterations in embryo development (*abi3*). Gene redundancy may explain the reduced number of mutants detected. Many Arabidopsis genes are in tandem arrays or segmental duplications [43]. We examined how many of the genes in our selection were part of gene tandem arrays or duplicated in different parts of the genome (Table 1). 11 of the selected genes (22%) are duplicated, which is higher than that observed in the whole genome (17%) ( $p$ -value = 0.33 in Fisher's exact test).

#### **Patterns of gene expression during silique and seed development**

The patterns of expression during seed development were investigated for each of the selected genes. Expression data was obtained from the Digital Northern tool in Genevestigator [24], corresponding to microarray hybridization of Affymetrix ATH1GeneChip® microarrays using labelled cDNAs of siliques and seeds at different stages of development, from mid-globular to green cotyledon embryos [9]. We used SOTA analysis in the TMEV 3.1 analysis package to identify expression patterns during silique and seed

development (Figure 4). From this analysis, we can distinguish four major patterns of expression (Table 1):

Group I: higher expression at early seed development. Genes that reach the maximum level of expression between late torpedo and early walking-stick embryo stages. This group includes five genes: At5g09640, encoding a serine carboxypeptidase, At5g49190, encoding a sucrose synthase, At2g34700, encoding a proline rich glycoprotein, and two genes encoding germin-like proteins (At3g04170 and At3g04190).

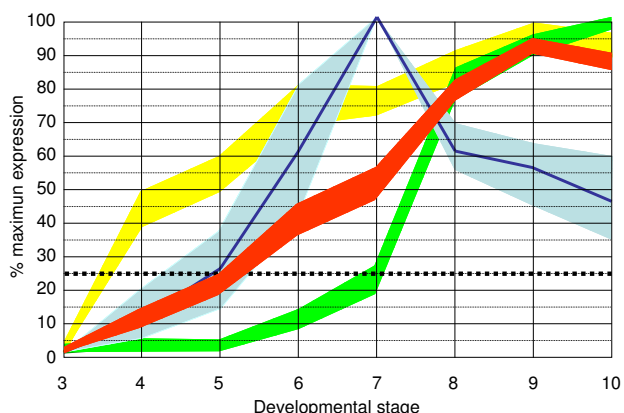
Group II: higher expression at mid seed development or later. The expression increases progressively, reaching the maximum level at the early cotyledon stage or later. In turn, SOTA analysis divided this class into three groups that can be distinguished by the stage at which their transcription level is higher than 25% of the maximum:

- IIa. Very early expression. The expression increases to more than 25% of the maximum before the early embryo stage. Four genes are included in this group. At5g48100, encoding a laccase, At4g36700, encoding a globulin-like protein, At4g37050, encoding a patatin-like protein, and At1g62060, encoding a protein of unknown function.
- IIb. Early expression. The expression increases to more than 25% of the maximum between the early heart and late torpedo stages. This group has 23 genes and includes the majority of the "nutrient reserve" genes.
- IIc. Mid stage expression. The expression increases to more than 25% of the maximum later than the late torpedo stage. It includes 17 genes of diverse functions.

#### **Conclusion**

Despite the technical problems associated with the relatively reduced number of Arabidopsis ESTs available, we have demonstrated here that the combination of EST profiling with microarray-based *in silico* selection may be a quick and cheap first step in the identification of Arabidopsis genes specifically expressed in certain organs, or in response to certain environmental stimuli. The same method could be applied to several other plant species in which EST sequences are available from several different organs and under different conditions (maize, wheat, rice, barley soybean, loblolly pine, etc). However, microarray data available for species other than Arabidopsis are very limited and less openly accessible, severely limiting the applicability of our two-step selection approach. An increase in EST sequencing, using more specific libraries, and in the contents of public microarray databases will greatly contribute to the efficiency of the method in plants.





**Figure 4**  
**Expression profiles during seed development showing four different patterns of expression in the subtracted genes.** Expression data are based on the microarray results [9]. Blue, pattern I; yellow, pattern IIa; red, pattern IIb; green, pattern IIc. Solid lines correspond to average expression and shaded areas to the standard errors. Developmental stages: 3, siliques with embryos at the mid-globular to early heart embryo stage; 4, siliques with embryos at the early to late heart-embryo stage; 5, siliques with embryos at the late heart to mid torpedo stages; 6, seeds with embryos at the late torpedo stage; 7, seeds with embryos at the late torpedo to early walking-stick stage; 8, seeds with embryos at the walking-stick to early curled-cotyledon stages; 9, seeds with embryos at the curled-cotyledon to early green-cotyledon stages; 10, seeds with embryos at the green cotyledon stage. The dotted line corresponds to 25% of the maximum expression.

**Methods**

**Plant material**

*Arabidopsis thaliana* Col-0 plants were grown in soil, in growth chambers, at 22 °C, with 18 h day. Plants used for root RNA extractions were grown on 0.8% (w/v) MS basal salt mixture agar plates in growth chambers, at 22 °C, with 18 h day.

**cDNA library construction and tag sequencing of expressed sequences**

Total RNA was extracted from frozen seeds as previously described [44] and treated with RNase-free DNaseI (Promega). Double stranded cDNA was built using the SMART cDNA Library Construction Kit (Clontech) according to the manufacturer's instructions, and introduced into the pCRII-TOPO (Invitrogen) vector for sequencing using the TOPO TA Cloning kit (Invitrogen).

For sequencing, DNA was amplified using PCR primers specific for the plasmid vector (5'-GTCACGACGTTGT-TAAACGACGGC-3' and 5'-GGAAACAGCTATGACCAT-GATTACG-3') and sequencing was carried out using a 5' specific primer (5'-GTATCAACGCAGAGTCG-3') and BigDye Terminator (Applied Biosystems) technology according to the manufacturer's instructions, in an ABI PRISM 3700 (Applied Biosystems). Cloning vector sequences were masked, and low quality and short (<190 bp) sequences removed. Homology searches for function assignment were performed using the BLASTN program in the Arabidopsis Information Resource (TAIR) [28]. EST sequences were deposited in the GeneBank database under the Accession numbers AM111128-AM111305.

**In Silico Subtraction**

Newly sequenced expressed sequence tags and 10,854 EST sequences of three libraries from immature *Arabidopsis* seeds (5564, 5576 and #C6I in TIGR Arabidopsis Gene Index [29]) were used as the initial source of immature seed sequences. *In silico* subtraction was done using a second set of EST libraries that did not contain immature seed sequences (50,992 ESTs from 78 libraries). Comparisons were based on the tentative gene contigs classification in the TIGR Arabidopsis database [29]. Libraries constructed from mixed tissues which could include immature seeds, such as immature siliques, were not considered for the subtraction. Subtraction was done by comparing the lists of genes that are represented in "immature seed" EST libraries with the list of genes represented by in "other organ" EST libraries.

**Table 3: Primers used for RT-PCR analysis**

Gene (Atg)	Forward primer	Reverse primer
At5g09640	GACACACCAAACATCAGAACCG	CTACTCATCATCCAAGGTCTCC
At5g22470	TATGCTCTCTCCGGTTCCTGG	ATGGAACCAACCGTCCACAAGG
At5g45690	ACGATTGCGACTCCTCTAAACC	GAACGGAGCCAATTTCTGCATC
At1g67100	GCTCATGAACCTCCTCAACACC	CCCGATCCAAGTCTTTGGTTCC
At3g60730	TCAAGCTGTGGCGTTGAGAGTG	GGTAAACGGAGAAGCCTCTTCC
At3g12203	GGCACTGATCTCTGATGAACAC	TTCTGAACCATCCATGGTCTCC
At1g71691	GCTTGTTCTTCATCGGAATGGG	TACGACAAGGCGTTTCAAAGGG
At2g43260	TTCCGGCTTGAACCATAACTGC	TGAACCACCTTTTCTGCCTTCG
At1g68380	TGTTTTATGGCCGCGTATTCC	TCCAAGTAAGCGTCTATTCCG
At4g14780	TCAAACCTCGCTCTTGATCTCGC	TTTACCACCTCCTTCATCTCC

A second selection step was based on the Arabidopsis Affymetrix GeneChip® data, available from the Meta-analyzer tool of the Genevestigator software [24-26]. Genes represented in the arrays with more than one probe were selected only when the results with all the probes passed the selection criteria.

### Gene Ontology

Functional characterization was performed according to the Gene Ontology (GO) Consortium through the Arabidopsis consortium information [28]. Fisher's exact test was performed using the MATFORSK, Norwegian Food Research Institute online facility [45,46].

### RT-PCR

Total RNAs were extracted from frozen organs of Arabidopsis as previously described [44] and treated with RNase-free DNaseI (Promega). Total pre-treated RNA (2 µg) was reverse transcribed with the Omniscript reverse transcriptase kit (Qiagen) using an oligo-dT primer. cDNAs were amplified with specific primers (Table 3), and controls, with non-reverse transcribed RNA, were also used to detect gDNA contamination. The actin gene was used as a control for RNA loading. PCR reactions were performed using 0.2 mM of each dNTP, 360 µg/ml BSA and 1 pmol µL<sup>-1</sup> of each primer in a final volume of 50 µL. The reaction mixtures were heated to 95°C for 5 min, followed by 28 cycles of 94°C for 30 sec, 55°C for 30 sec, and 72°C for 90 sec. Reactions were completed by incubating at 72°C for 10 min. The amounts of template cDNA and the number of PCR cycles were determined for each gene to ensure that amplification occurred in the linear range and allowed for good comparison of the amplified products. At least two independent analyses were carried out on the different RNA samples. Reactions were performed in a Minicycler (MJ Research, Waltham, MA) thermal cycler.

### In situ hybridization

The protocol for *in situ* hybridization was done as previously described [47] except for the labelling of the probes and the detection of the signal. Probes were synthesized and labelled using the Boehringer digoxigenin system, and detected using the BM purple AP substrate (Boehringer). The probe was synthesized from the product of PCR amplification cloned into the pCRII-TOPO vector (Invitrogen).

### Gene distribution in tandem arrays and mutants

The presence of the selected genes in tandem arrays was based on previously described data [43]. Genes whose loss-of-function give an embryo mutant phenotype were determined according to data previously collected [6,7].

### Expression cluster analysis

For expression cluster analysis, we used the TIGR Multi Experiment Viewer (TMEV) software [48]. Original data was obtained from the Genevestigator tool [24-26] and correspond to a microarray analysis of silique and seed development [9].

### Authors' contributions

CB carried out the experimental molecular genetic studies. Database searches and analyses were performed by CMV and CB. PP supervised the study and wrote the manuscript jointly with CMV and CB.

### Additional material

#### Additional file 1

*Libraries used in the subtraction process step 1* Data obtained from the TIGR Arabidopsis Gene Index [http://www.tigr.org/tigr-scripts/tgi/T\\_index.cgi?species=arab](http://www.tigr.org/tigr-scripts/tgi/T_index.cgi?species=arab).

Click here for file

[<http://www.biomedcentral.com/content/supplementary/1471-2164-7-38-S1.doc>]

#### Additional file 2

*Genes selected by EST subtraction* Genes having corresponding EST sequences in immature seed libraries and not in libraries of other tissues.

Click here for file

[<http://www.biomedcentral.com/content/supplementary/1471-2164-7-38-S2.doc>]

### Acknowledgements

This work was carried out thanks to grants BIO2001-1721 and BIO2004-01577 from the Plan Nacional de Investigación Científica y Técnica and a grant from the program MAZE, European Union, and within the framework of Centre de Referència de Biotecnologia de la Generalitat de Catalunya. C.B. was the recipient of a fellowship from the Universitat Autònoma de Barcelona – Fundació Presidente Allende. C.M.V. is the recipient of a "Ramon y Cajal" contract from the Spanish Ministry of Science.

### References

- Olsen OA: **Endosperm development: cellularization and cell fate specification.** *Annu Rev Plant Physiol Plant Mol Biol* 2001, **52**:233-267.
- Willemsen V, Scheres B: **Mechanisms of pattern formation in plant embryogenesis.** *Annu Rev Genet* 2004, **38**:587-614.
- McElver J, Tzafirir I, Aux G, Rogers R, Ashby C, Smith K, Thomas C, Schetter A, Zhou Q, Cushman MA, Tossberg J, Nickle T, Levin JZ, Law M, Meinke D, Patton D: **Insertional mutagenesis of genes required for seed development in Arabidopsis thaliana.** *Genetics* 2001, **159**:1751-1763.
- Chaudhury AM, Koltunow A, Payne T, Luo M, Tucker MR, Dennis ES, Peacock WJ: **Control of early seed development.** *Annu Rev Cell Dev Biol* 2001, **17**:677-699.
- Meinke DW, Meinke LK, Showalter TC, Schissel AM, Mueller LA, Tzafirir I: **A sequence-based map of Arabidopsis genes with mutant phenotypes.** *Plant Physiol* 2003, **131**:409-418.
- Tzafirir I, Pena-Muralla R, Dickerman A, Berg M, Rogers R, Hutchens S, Sweeney TC, McElver J, Aux G, Patton D, Meinke D: **Identification of genes required for embryo development in Arabidopsis.** *Plant Physiol* 2004, **135**:1206-1220.
- SeedGenes Project** [<http://www.seedgenes.org/>]

8. Ma L, Sun N, Liu X, Jiao Y, Zhao H, Deng XW: **Organ-specific Expression of Arabidopsis Genome during development.** *Plant Physiol* 2005, **138**:80-91.
9. Schmid M, Davison TS, Henz SR, Pape UJ, Demar M, Vingron M, Scholkopf B, Weigel D, Lohmann JU: **A gene expression map of Arabidopsis thaliana development.** *Nat Genet* 2005, **37**:501-506.
10. Bernstein SL, Borst DE, Neuder ME, Wong P: **Characterization of the human fovea cDNA library and regional differential gene expression in the human retina.** *Genomics* 1996, **32**:301-308.
11. Vasmatzis G, Essand M, Brinkmann U, Lee B, Pastan I: **Discovery of three genes specifically expressed in human prostate by expressed sequence tag database analysis.** *Proc Natl Acad Sci USA* 1998, **95**:300-304.
12. Itoh K, Okubo K, Utiyama H, Hirano T, Yoshii J, Matsubara K: **Expression profile of active genes in granulocytes.** *Blood* 1998, **92**:1432-1441.
13. Bortoluzzi S, d'Alessi F, Romualdi C, Danieli GA: **The human adult skeletal muscle transcriptional profile reconstructed by a novel computational approach.** *Genome Res* 2000, **10**:344-349.
14. Huminiecki L, Bicknell R: **In silico cloning of novel endothelial-specific genes.** *Genome Res* 2000, **10**:1796-1806.
15. Miner D, Rajkovic A: **Identification of expressed sequence tags preferentially expressed in human placentas by in silico subtraction.** *Prenat Diagn* 2003, **23**:410-419.
16. Baranova AV, Lobashev AV, Ivanov DV, Krukovskaya LL, Yankovsky NK, Kozlov AP: **In silico screening for tumour-specific expressed sequences in human genome.** *FEBS Lett* 2001, **508**:143-148.
17. **NCBI Expressed Sequence Tags database** [<http://www.ncbi.nlm.nih.gov/dbEST/>]
18. Ogihara Y, Mochida K, Nemoto Y, Murai K, Yamazaki Y, Shin-I T, Kohara Y: **Correlated clustering and virtual display of gene expression patterns in the wheat life cycle by large-scale statistical analyses of expressed sequence tags.** *Plant J* 2003, **33**:1001-1011.
19. Casu RE, Dimmock CM, Chapman SC, Grof CP, McIntyre CL, Bonnett GD, Manners JM: **Identification of differentially expressed transcripts from maturing stem of sugarcane by in silico analysis of stem expressed sequence tags and gene expression profiling.** *Plant Mol Biol* 2004, **54**:503-517.
20. Redman JC, Haas BJ, Tanimoto G, Town CD: **Development and evaluation of an Arabidopsis whole genome Affymetrix probe array.** *Plant J* 2004, **38**:545-561.
21. Honys D, Twell D: **Transcriptome analysis of haploid male gametophyte development in Arabidopsis.** *Genome Biol* 2004, **5**:R85.
22. Lloyd JC, Zakhleniuk OV: **Responses of primary and secondary metabolism to sugar accumulation revealed by microarray expression analysis of the Arabidopsis mutant, pho3.** *J Exp Bot* 2004, **55**:1221-1230.
23. Menges M, de Jager SM, Gruissem W, Murray JA: **Global analysis of the core cell cycle regulators of Arabidopsis identifies novel genes, reveals multiple and highly specific profiles of expression and provides a coherent model for plant cell cycle control.** *Plant J* 2005, **41**:546-566.
24. **Genevestigator** [<http://www.genevestigator.ethz.ch/>]
25. Zimmermann P, Hirsch-Hoffmann M, Hennig L, Gruissem W: **GENEVESTIGATOR. Arabidopsis Microarray Database and Analysis Toolbox.** *Plant Physiol* 2004, **136**:2621-2632.
26. Zimmermann P, Hennig L, Gruissem W: **Gene-expression analysis and network discovery using Genevestigator.** *Trends Plant Sci* 2005, **10**:407-409.
27. White JA, Todd J, Newman T, Focks N, Girke T, Martínez de Ilárduya O, Jaworski JG, Ohlrogge JB, Benning C: **A New Set of Arabidopsis Expressed Sequence Tags from Developing Seeds. The Metabolic Pathway from Carbohydrates to Seed Oil.** *Plant Physiol* 2000, **124**:1582-1594.
28. **The Arabidopsis Information Resource, TAIR** [<http://www.arabidopsis.org/>]
29. **TIGR Arabidopsis Gene Index** [[http://www.tigr.org/tigr-scripts/tgi/T\\_index.cgi?species=arab](http://www.tigr.org/tigr-scripts/tgi/T_index.cgi?species=arab)]
30. Giraudat J, Hauge BM, Valon C, Smalle J, Parcy F, Goodman HM: **Isolation of the Arabidopsis ABI3 gene by positional cloning.** *The Plant Cell* 1992, **4**:1251-1261.
31. Haslekas C, Stacy RA, Nygaard V, Culianez-Macia FA, Aalen RB: **The expression of a peroxiredoxin antioxidant gene, AtPer1, in Arabidopsis thaliana is seed-specific and related to dormancy.** *Plant Mol Biol* 1998, **36**:833-845.
32. Dong J, Keller WA, Yan W, Georges F: **Gene expression at early stages of Brassica napus seed development as revealed by transcript profiling of seed-abundant cDNAs.** *Planta* 2004, **218**:483-491.
33. Nuccio ML, Thomas TL: **ATSI and ATS3: two novel embryo-specific genes in Arabidopsis thaliana.** *Plant Mol Biol* 1999, **39**:1153-1163.
34. Vicient CM, Hull G, Guillemot J, Devic M, Delseny M: **Differential expression of the Arabidopsis genes coding for Em-like proteins.** *J Exp Bot* 2000, **51**:1211-1220.
35. Rudd S: **Expressed sequence tags: alternative or complement to whole genome sequences?** *Trends Plant Sci* 2003, **8**:321-329.
36. Berardini TZ, Mundodi S, Reiser L, Huala E, Garcia-Hernandez M, Zhang P, Mueller LA, Yoon J, Doyle A, Lander G, Moseyko N, Yoo D, Xu I, Zoeckler B, Montoya M, Miller N, Weems D, Rhee SY: **Functional Annotation of the Arabidopsis Genome Using Controlled Vocabularies.** *Plant Physiology* 2004, **135**:745-755.
37. Vicente-Carbajosa J, Carbonero P: **Seed maturation: developing an intrusive phase to accomplish a quiescent state.** *Int J Dev Biol* 2005, **49**:645-651.
38. Borisjuk L, Rolletschek H, Radchuk R, Weschke W, Wobus U, Weber H: **Seed development and differentiation: a role for metabolic regulation.** *Plant Biol* 2004, **6**:375-386.
39. Souter M, Lindsey K: **Polarity and signalling in plant embryogenesis.** *J Exp Bot* 2000, **51**:971-983.
40. Jose-Estanyol M, Ruiz-Avila L, Puigdomenech P: **A maize embryo-specific gene encodes a proline-rich and hydrophobic protein.** *The Plant Cell* 1992, **4**:413-423.
41. Bai S, Chen L, Yund MA, Sung ZR: **Mechanisms of plant embryo development.** *Curr Top Dev Biol* 2000, **50**:61-88.
42. Lehti-Shiu MD, Adamczyk BJ, Fernandez DE: **Expression of MADS-box genes during the embryonic phase in Arabidopsis.** *Plant Mol Biol* 2005, **58**:89-107.
43. Haberer G, Hindemitt T, Meyers BC, Mayer KF: **Transcriptional similarities, dissimilarities, and conservation of cis-elements in duplicated genes of Arabidopsis.** *Plant Physiol* 2004, **136**:3009-3022.
44. Vicient CM, Delseny M: **Isolation of total RNA from Arabidopsis thaliana seeds.** *Anal Biochem* 1999, **268**:412-413.
45. **MATFORSK (Norwegian Food Research Institute) Øyvind Langrudonline Fisher's exact test facility** [<http://www.matforsk.no/ola/fisher.htm>]
46. Agresti A: **A Survey of Exact Inference for Contingency Tables.** *Statistical Science* 1992, **7**:131-153.
47. Cox KH, DeLeon DV, Angerer LM, Angerer RC: **Detection of mRNAs in sea urchin embryos by in situ hybridization using asymmetric RNA probes.** *Dev Biol* 1984, **101**:485-502.
48. **TIGR Multi Experiment Viewer (TMEV) software** [<http://www.tigr.org/software/>]
49. Bowman JL: *Arabidopsis: an Atlas of Morphology and Development* Berlin & New York: Springer-Verlag; 1993.

Publish with **BioMed Central** and every scientist can read your work free of charge

"BioMed Central will be the most significant development for disseminating the results of biomedical research in our lifetime."

Sir Paul Nurse, Cancer Research UK

Your research papers will be:

- available free of charge to the entire biomedical community
- peer reviewed and published immediately upon acceptance
- cited in PubMed and archived on PubMed Central
- yours — you keep the copyright

Submit your manuscript here:  
[http://www.biomedcentral.com/info/publishing\\_adv.asp](http://www.biomedcentral.com/info/publishing_adv.asp)





## **APÉNDICE II**

**Genes seleccionados por sustracción de ESTs**

Lista completa de los genes para los cuales se han secuenciado ESTs a partir de genotecas de cDNA de semilla inmadura y no de otros órganos.

Código AGI o N. Acc. GeneBank	Código TIGR	Definición
At1g01225	TC267364	NC domain-containing protein-related
At1g02740	TC278236	MRG family protein
At1g02770	TC275927	Unknown
At1g03103	TC268770	Protease inhibitor/seed storage/lipid transfer protein (LTP) family protein
At1g03106	TC273344	Unknown
At1g03210	TC276036	Similar to phenazine biosynthesis phzc/phzf family protein
At1g03560	TC255717	Pentatricopeptide (PPR) repeat-containing protein
At1g03790	TC274104	Zinc finger (CCCH-type) family protein
At1g03890	TC251613	Cruciferin 12S seed storage protein
At1g03920	TC275977	Protein kinase MK6
At1g04160	TC277571	Myosin heavy chain MYA2
At1g04880	TC257963	Glutathione S-transferase GST16-like
At1g05280	TC278832	Fringe-related protein
At1g05450	TC255931	Lipid-transfer protein putative
At1g06410	TC263622	Glycosyl transferase family 20 protein
At1g06450	TC274674	CCR4-NOT transcription complex protein
At1g06820	TC256058	Carotenoid isomerase
At1g07705	TC266925	VIP2 protein
At1g07950	TC254374	Surfeit locus protein 5 family protein
At1g08060	TC263990	MOM1
At1g08420	TC267099	Serine/threonine phosphoesterase family protein
At1g08810	TC274237	MYB family transcription factor
At1g09155	TC276278	SKP1 interacting partner 3-related
At1g09190	TC257368	Pentatricopeptide (PPR) repeat-containing protein
At1g09380	TC263149	Putative nodulin protein N21
At1g09490	TC263673	Alcohol dehydrogenase
At1g09550	TC268517	Pectinacetyltransferase
At1g09580	TC267661	Transmembrane protein Tmp21 precursor
At1g09790	TC267949	COBRA-like protein 6 precursor
At1g09950	TC274829	Transcription factor-related
At1g10120	TC268790	Basic helix-loop-helix (bhlh) family protein
At1g10210	TC254539	Mitogen-activated protein kinase homolog 1 (MAP kinase 1) (atmpk1)
At1g10520	TC267641	DNA polymerase lambda
At1g10640	TC269021	Polygalacturonase
At1g10750	TC275636	Putative carboxyl-terminal peptidase
At1g11170	TC253269	Unknown
At1g11190	TC262545	Bifunctional nuclease (BFN1)
At1g11590	TC255226	Putative pectin methyltransferase
At1g11720	TC257480	Starch synthase
At1g11960	TC257361	Early-responsive to dehydration protein-related
At1g12230	TC278590	Transaldolase-like protein
At1g12550	TC265059	Oxidoreductase family protein
At1g12805	TC257360	Unknown
At1g14260	TC265415	Zinc finger (C3HC4-type RING finger) family protein
At1g14580	TC259683	Putative zinc finger protein
At1g14950	TC252695	Major latex protein type1
At1g14970	TC257070	Unknown
At1g15150	TC254873	MATE efflux family protein
At1g15200	TC254630	Protein-protein interaction regulator family protein
At1g15330	TC273248	CBS domain-containing protein
At1g15510	TC268319	Pentatricopeptide (PPR) repeat-containing protein
At1g16040	TC277578	Phosphatidylinositol-glycan biosynthesis class F protein
At1g16980	TC268239	Alpha-trehalose-phosphate synthase
At1g17060	TC256663	Cytochrome P450-like protein
At1g17380	TC268739	Unknown
At1g17650	TC255116	6-phosphogluconate dehydrogenase NAD-binding domain-containing protein
At1g18950	TC258930	Aminoacyl-tRNA synthetase
At1g20260	TC270864	Vacuolar ATP synthase subunit B
At1g20410	TC264876	Unknown
At1g20500	TC275366	4-coumarate-coa ligase-like protein
At1g21710	TC255556	8-oxoguanine DNA glycosylase
At1g21730	TC256127	Kinesin-related protein
At1g21740	TC278038	Unknown
At1g22020	TC264213	Glycine hydroxymethyltransferase
At1g23980	TC256248	RING zinc finger protein-like
At1g24430	TC256521	Similar to deacetylindoline 4-O-acetyltransferase
At1g25054	TC268745	UDP-3-O-acyl N-acetylglucosamine deacetylase
At1g25470	TC255477	Subfamily B-6 of ERF/AP2 transcription factor family.
At1g26090	TC257571	Unknown
At1g26370	TC275409	RNA helicase
At1g26570	TC264346	UDP-glucose dehydrogenase
At1g26680	TC257022	Transcriptional factor B3 family protein
At1g27040	TC263230	Putative nitrate transporter
At1g27590	TC276948	Unknown
At1g27760	TC271419	Unknown
At1g27960	TC279820	Unknown

At1g28500	TC258852	Unknown
At1g28540	TC275708	Unknown
At1g29400	TC271188	Probable RNA-binding protein
At1g29750	TC264672	Similar to leucine-rich repeat transmembrane protein kinase
At1g30550	TC259577	Unknown
At1g31470	TC257109	Nodulin-related
At1g32260	TC275047	Unknown
At1g33050	TC271867	Unknown
At1g33420	TC265403	Phd finger family protein
At1g34210	TC256464	Somatic embryogenesis receptor-like kinase 2
At1g34360	TC256692	Translation initiation factor 3 (if-3) family protein
At1g35140	TC264710	Probable phosphate-induced (phi-1) protein
At1g35510	TC278500	Unknown
At1g36160	TC272263	Similar to acetyl-coa carboxylase 2
At1g43675	TC277623	Surfeit locus protein 5 family protein
At1g44110	TC256992	Mitotic cyclin a2-type
At1g45180	TC276096	ZINC FINGER protein
At1g48130	TC262619	Peroxiredoxin
At1g48660	TC258441	Auxin-responsive GH3 family protein similar to auxin-responsive GH3 product
At1g50950	TC268815	Thioredoxin-related
At1g54080	TC261325	Oligouridylylate binding protein putative partial
At1g54150	TC274921	Zinc finger
At1g54510	TC265231	Protein kinase family protein
At1g55080	TC274231	Unknown
At1g55270	TC272209	Kelch repeat-containing f-box family protein
At1g55940	TC277513	Cytochrome p450
At1g56180	TC264385	Unknown
At1g60987	-	Cysteine rich protein
At1g61690	TC274482	Tetratricopeptide repeat (TPR)-containing protein
At1g61720	TC264032	Dihydroflavonol 4-reductase
At1g61810	TC275992	Glycosyl hydrolase family 1 protein
At1g61820	TC263171	Glycosyl hydrolase family 1 protein
At1g61860	TC257044	Protein kinase
At1g62000	TC261464	Unknown
At1g62060	TC261285	Unknown
At1g62080	TC261521	Unknown
At1g62220	TC254029	Unknown
At1g62225	TC271828	Unknown
At1g62860	TC278466	Protein kinase
At1g62880	TC273537	Cornichon family protein
At1g63020	TC266722	RNA polymerase IIA largest subunit putative
At1g63140	TC255296	Caffeic O-methyltransferase
At1g63140	TC255297	O-methyltransferase
At1g63370	TC274790	Flavin-containing monooxygenase family protein
At1g63650	TC264217	Transcription factor EGL1 (Basic helix-loop-helix protein 2)
At1g64490	TC266621	Unknown
At1g64580	TC274426	Pentatricopeptide (ppr) repeat-containing protein
At1g65090	TC264499	Unknown
At1g65880	TC265191	Amp-dependent synthetase and ligase family protein
At1g66460	TC275597	Protein kinase-like protein
At1g67100	TC272916	Similar to seed specific protein Bn15D17A
At1g67420	TC268852	Peptidase family-like protein
At1g68120	TC267291	Unknown
At1g68170	TC278190	Mtn21-like protein
At1g68380	TC274078	Unknown
At1g68460	TC266602	Cytokinin synthase
At1g69040	TC253973	Act domain containing protein
At1g69570	TC264709	Dof-type zinc finger domain-containing protein
At1g69670	TC256753	Cullin 3B
At1g70895	TC267074	Clavata3/esr-related 17
At1g71120	TC257705	Gdsl-motif lipase/hydrolase family protein
At1g71250	TC255820	GDSL-motif lipase/hydrolase/hydrolase
At1g71691	TC273517	GDSL-motif lipase/hydrolase/hydrolase
At1g71950	TC278289	Unknown
At1g72190	TC256049	Phosphoglycerate dehydrogenase
At1g72270	TC276544	Unknown
At1g72560	TC257501	Trna export mediator exportin-t
At1g72600	TC276178	Hydroxyproline-rich glycoprotein family protein
At1g72670	TC276961	Calmodulin-binding family protein
At1g72830	TC253699	CCAAT-binding factor B subunit-like protein
At1g73190	TC252299	Tonoplast intrinsic protein 3.1
At1g73290	TC257943	Serine carboxypeptidase
At1g73550	TC267238	Lipid transfer protein
At1g75020	TC258136	Phospholipid/glycerol acyltransferase family protein
At1g75860	TC276433	Unknown
At1g76110	TC254387	High mobility group (hmg1/2)
At1g76290	TC277723	Amp-dependent synthetase and ligase family protein
At1g76630	TC267677	Tetratricopeptide repeat (tpr)-containing protein
At1g76880	TC274329	DNA-binding protein DF1
At1g77300	TC269040	SET domain-containing protein
At1g78390	TC266589	Putative 9-cis-epoxycarotenoid dioxygenase
At1g78540	TC259602	Transcription factor-related

At1g79430	TC254186	Myb DNA binding protein
At1g79690	TC276846	Mutl/nudix family protein
At1g79840	TC264438	Homeobox protein GLABRA2
At1g79950	TC276691	Helicase-related
At1g80090	TC278341	CBS domain-containing protein
At1g80370	TC258460	Cyclin
At2g01130	TC265554	Helicase domain-containing protein
At2g01610	TC266479	Invertase/pectin methylesterase inhibitor family protein
At2g01750	TC275097	Microtubule associated protein
At2g02120	TC264415	Plant defensin-fusion protein
At2g02490	-	Proline-rich family protein
At2g03390	TC258931	UvrB/uvrC motif-containing protein
At2g03780	TC256169	Translin family protein
At2g04039	TC275315	Unknown
At2g05760	TC276479	Xanthine/uracil permease family protein
At2g07170	TC277832	Unknown
At2g07718	TC257239	Cytochrome b
At2g14680	TC267562	Myosin heavy chain-related
At2g17300	TC267464	Unknown
At2g18160	TC261643	Bzip transcription factor family protein
At2g18850	TC275254	Set domain-containing protein
At2g18915	TC265884	E3 ubiquitin ligase scf complex f-box subunit
At2g20310	TC267539	Unknown
At2g20440	TC266613	Rabgap/tbc domain-containing protein
At2g20770	TC274172	Lanthionine synthetase c-like family protein
At2g21185	TC263681	Unknown
At2g21260	TC273647	NADPH dependent mannose 6-phosphate reductase
At2g22030	TC267317	Kelch repeat-containing f-box family protein
At2g22560	TC277813	Kinase interacting protein-related
At2g22570	TC278498	Isochorismatase hydrolase family protein
At2g22910	TC256168	Putative amino acid acetyltransferase
At2g23140	TC256913	Armadillo/beta-catenin repeat family protein
At2g23550	TC266043	Acetone-cyanohydrin lyase
At2g23560	TC278169	Acetone-cyanohydrin lyase
At2g23580	TC273923	Acetone-cyanohydrin lyase
At2g24640	TC255345	Ubiquitin carboxyl terminal hydrolase
At2g25940	TC254259	Vacuolar processing enzyme alpha-isozyme precursor (Alpha-VPE)
At2g26580	TC274658	Plant-specific transcription factor yabby family protein
At2g27240	TC268181	Unknown
At2g27775	TC279724	Unknown
At2g28240	TC265858	Putative hydroxyproline-rich glycoprotein
At2g28420	TC274367	Lactoylglutathione lyase family protein / glyoxalase I family protein
At2g28650	TC274204	Exocyst subunit exo70 family protein
At2g30300	TC277331	Nodulin-related
At2g30350	TC267880	Endo/excinuclease amino terminal domain
At2g30615	TC256068	Unknown
At2g30680	TC268036	Similar to glycosyl transferase family 48 protein
At2g30800	TC264682	ATP-dependent RNA helicase A
At2g30942	TC269210	Unknown
At2g32330	TC259172	Unknown
At2g32670	TC265205	Synaptobrevin family protein
At2g33240	TC268808	Myosin
At2g33520	TC275699	Unknown
At2g33580	TC259106	Peptidoglycan-binding lysm domain-containing protein kinase
At2g33690	TC276402	Late embryogenesis abundant protein
At2g34700	TC264743	Extensin family protein
At2g35530	TC268009	Bzip transcription factor family protein
At2g36400	TC274405	Transcription activator grl3
At2g36760	TC277238	Udp-glucuronosyl/udp-glucosyl transferase family protein
At2g37025	TC256715	Pathogen-responsive dna-binding protein-related
At2g37090	TC272851	Glycosyltransferase
At2g38090	TC254271	MYB transcription factor
At2g38590	TC267548	F-box family protein
At2g38740	TC273682	Haloacid dehalogenase-like hydrolase family protein
At2g38920	TC275414	Spx (syg1/pho81/xpr1) domain-containing protein
At2g39560	TC256033	Unknown
At2g39680	TC265171	Trans-acting sirna primary transcript
At2g40150	TC265866	Unknown
At2g41070	TC252623	Bzip protein DPBF4
At2g41140	TC276380	Calcium-dependent protein kinase
At2g41540	TC262537	Glycerol-3-phosphate dehydrogenase
At2g41700	TC254835	ATP-binding cassette transporter atabca1
At2g41880	TC275970	Guanylate kinase 1
At2g42110	TC276810	Unknown
At2g42170	TC276214	Actin 2
At2g42950	TC265979	Unknown
At2g43190	TC255068	Ribonuclease p family protein
At2g43260	TC254993	F-box family protein
At2g43400	TC262174	Electron transfer flavoprotein-ubiquinone oxidoreductase family protein
At2g43880	TC276244	Polygalacturonase
At2g44470	TC254864	Putative beta-glucosidase
At2g45250	TC276816	Unknown

At2g45420	TC265771	LOB domain protein 18
At2g45900	TC277822	Unknown
At2g46550	TC262497	Unknown
At2g46960	TC252957	Cytochrome P450-like protein
At2g47120	TC278076	Short-chain dehydrogenase/reductase (sdr) family protein
At2g47750	TC273105	Auxin-regulated protein GH3 homolog
At3g01160	TC265914	Unknown
At3g01570	TC251579	Oleosin
At3g02890	TC257489	Phd finger protein-related
At3g03020	TC277593	Unknown
At3g03240	TC258420	Esterase/lipase/thioesterase family protein
At3g03300	TC266924	Dead/deah box helicase carpel factory-related
At3g04170	TC276999	Germin-like protein subfamily 1 member 3 precursor
At3g04180	TC255981	Germin-like protein subfamily 1 member 4 precursor
At3g04190	TC267823	Germin-like protein subfamily 1 member 5 precursor
At3g04660	TC258810	F-box family protein
At3g04690	TC268418	Protein kinase family protein
At3g04960	TC257772	Unknown
At3g05400	TC253408	Putative sugar transporter
At3g05800	TC257173	Unknown
At3g07200	TC273968	ZINC FINGER family protein
At3g07730	TC257774	Unknown
At3g08900	TC265817	UDP-glucose:protein transglucosylase-like protein
At3g08910	TC275769	DNAJ heat shock protein
At3g08970	TC255587	Putative dnaj protein
At3g10400	TC275282	Rna recognition motif (rrm)-containing protein
At3g11020	TC274803	Dreb subfamily a-2 of erf/ap2 transcription factor family
At3g11180	TC264077	Leucoanthocyanidin dioxygenase-like protein
At3g11590	TC274447	Unknown
At3g11650	TC257781	Harpin-induced family protein
At3g12203	TC264046	Serine carboxypeptidase
At3g12550	TC277084	Xh/xs domain-containing protein
At3g12960	TC267267	Expressed protein similar to seed maturation protein PM28
At3g13030	TC264649	Hat dimerisation domain-containing protein
At3g13640	TC274297	Rnase I inhibitor protein
At3g14130	TC255339	Glycolate oxidase
At3g14630	TC266176	Cytochrome p450
At3g15050	TC258786	Calmodulin-binding family protein
At3g15510	TC263196	Atnac2
At3g15680	TC275230	Zinc finger (ran-binding) family protein
At3g17030	TC278454	Unknown
At3g17680	TC255787	Unknown
At3g18570	TC255057	Oleosin
At3g18950	TC265769	Transducin family protein
At3g18970	TC267906	Pentatricopeptide (ppr) repeat-containing protein
At3g19260	TC272922	Longevity-assurance (lag1) family protein
At3g19500	TC256280	Ethylene-responsive protein -related
At3g19870	TC275442	Unknown
At3g20420	TC256243	Ribonuclease iii family protein
At3g20520	TC273044	Probable glycerophosphoryl diester phosphodiesterase 2 precursor
At3g20680	TC274468	Unknown
At3g20810	TC274577	Transcription factor jumonji (jmc) domain-containing protein
At3g20840	TC277111	Similar to ovule development protein
At3g21100	TC255533	Rna recognition motif (rrm)-containing protein
At3g21400	TC274701	Unknown
At3g21730	TC276658	Dihydroneopterin aldolase family protein
At3g22180	TC266745	Zinc finger (dhhc type) family protein
At3g22220	TC274342	Hat dimerisation domain-containing protein
At3g22400	TC262490	Lipoxygenase
At3g23780	TC268543	DNA-directed RNA polymerase
At3g23900	TC256433	Rna recognition motif (rrm)-containing protein
At3g24220	TC257098	9-cis-epoxycarotenoid dioxygenase
At3g24650	TC273434	ABI3 protein
At3g25160	TC274836	Er lumen protein retaining receptor family protein
At3g25870	TC274449	Unknown
At3g25890	TC257333	Ethylene response factor
At3g26730	TC268833	Zinc finger (C3HC4-type RING finger) family protein
At3g26770	TC255764	Alcohol dehydrogenase-like protein
At3g26790	TC254343	Fusca3
At3g27260	TC266476	Dna-binding bromodomain-containing protein
At3g27660	TC275195	Oleosin
At3g27870	TC259551	Phospholipid-transporting atpase 8 (Aminophospholipid flippase 8)
At3g29390	TC255450	Hydroxyproline-rich glycoprotein
At3g30350	TC279515	Unknown
At3g30413	TC278248	Gypsy-like retrotransposon
At3g30436	TC258958	Gypsy-like retrotransposon
At3g42940	TC267081	Unknown
At3g44460	TC254735	Basic leucine zipper transcription factor
At3g44690	TC258394	Unknown
At3g44830	TC277330	Lecithin:cholesterol acyltransferase family protein
At3g45130	TC266003	Cycloartenol synthase
At3g45210	TC274871	Unknown

At3g46020	TC275270	RNA binding protein-like
At3g46720	TC277036	Glucuronosyl transferase-like protein
At3g47220	TC256796	Phosphoinositide-specific phospholipase c family protein
At3g48580	TC276359	Xyloglucan:xyloglucosyl transferase
At3g49140	TC256990	Unknown
At3g49640	TC268458	Nitrogen regulation family protein
At3g49740	TC267510	Pentatricopeptide (ppr) repeat-containing protein
At3g50410	TC256959	Dof-type zinc finger domain-containing protein
At3g50620	TC276937	Nodulation protein-related
At3g51370	TC276638	Protein phosphatase 2C
At3g52110	TC266567	Unknown
At3g52340	TC274220	Sucrose-phosphatase
At3g53030	TC264479	Protein kinase family protein
At3g53520	TC271113	DTDP-glucose 4-6-dehydratase-like protein
At3g54270	TC256776	Sucrose-phosphatase 3 (SPP3)
At3g54720	TC264539	Glutamate carboxypeptidase
At3g54860	TC263524	Vacuolar protein sorting protein
At3g54940	TC252346	Cysteine proteinase
At3g54970	TC274670	Unknown
At3g55020	TC271665	Rabgap/TBC domain-containing protein
At3g55090	TC268232	Abc transporter family protein
At3g55160	TC277915	Unknown
At3g56260	TC266131	Unknown
At3g56270	TC263812	Unknown
At3g57780	TC266632	Unknown
At3g58740	TC254249	Citrate synthase-like protein
At3g58780	TC256368	Agamous-like MADS box protein AGL1 (Protein Shatterproof 1)
At3g58790	TC272331	Glycosyl transferase family 8 protein
At3g59850	TC276255	Polygalacturonase-like protein
At3g60730	TC275601	Pectinesterase-like protein
At3g61040	TC256844	Cytochrome P450 monooxygenase-like protein
At3g61340	TC256774	F-box family protein
At3g61380	TC276237	Unknown
At3g61640	TC269184	Arabinogalactan-protein (AGP20)
At3g61690	TC273636	Unknown
At3g62730	TC270840	Desiccation-related protein
At3g63040	TC257442	Unknown
At4g00250	TC274705	Dna-binding storekeeper protein-related
At4g00540	TC274911	Myb family transcription factor
At4g00760	TC256981	Two-component responsive regulator family protein
At4g00790	TC257825	Unknown
At4g01650	TC272307	Unknown
At4g01897	TC259129	Unknown
At4g01970	TC265816	Raffinose synthase family protein
At4g02000	TC256985	Zinc finger protein
At4g02030	TC263409	Unknown
At4g02630	TC254789	Serine/threonine-specific protein kinase
At4g02740	TC256802	F-box family protein
At4g04155	TC269213	Unknown
At4g07960	TC272947	Glycosyl transferase family 2
At4g08330	TC273168	Unknown
At4g09180	TC274734	Basic helix-loop-helix (bhlh) family protein
At4g09500	TC274601	Glycosyltransferase family protein
At4g10850	TC274946	Nodulin mtn3 family protein
At4g12540	TC267418	Unknown
At4g12870	TC254854	Unknown
At4g13750	TC266982	Unknown
At4g14590	TC266891	Unknown
At4g14600	TC273938	Unknown
At4g14730	TC278401	Transmembrane protein-related
At4g14780	TC274223	Kinase like protein
At4g15020	TC265088	Unknown
At4g15120	TC255132	Vq motif-containing protein
At4g15396	TC266100	Cytochrome p450-related
At4g15417	TC268052	Ribonuclease iii family protein
At4g16620	TC257359	Integral membrane family protein
At4g17370	TC276649	Oxidoreductase family protein
At4g17895	TC275845	Ubiquitin-specific protease 20
At4g18910	TC273199	Aquaglyceroporin
At4g19160	TC271283	Unknown
At4g19500	TC254674	Resistance protein-like
At4g21060	TC277119	Galactosyltransferase family protein
At4g21380	TC255785	S-receptor kinase ARK3 precursor
At4g22100	TC259055	Glucosidase like protein
At4g22120	TC258222	Early-responsive to dehydration protein-related
At4g22290	TC254214	Ubiquitin carboxyl-terminal hydrolase family protein
At4g22390	TC257645	F-box family protein-related
At4g22560	TC275720	Unknown
At4g22820	TC271149	Zinc finger (an1-like) family protei
At4g22850	TC274318	Unknown
At4g23520	TC266928	Cysteine proteinase
At4g25140	TC274793	Oleosin

At4g25750	TC266771	Abc transporter family protein
At4g26400	TC262098	Zinc finger (c3hc4-type ring finger) family protein
At4g26420	TC265484	S-adenosyl-L-methionine:salicylic acid carboxyl methyltransferase-like protein
At4g27040	TC253797	SNF8 like protein
At4g27150	TC261110	2S seed storage protein 2 precursor
At4g27420	TC258290	Abc transporter family protein
At4g27460	TC265034	Cbs domain-containing protein
At4g28520	TC270748	12S cruciferin seed storage protein (CRU3)
At4g28530	TC254906	CUC2-like protein
At4g28760	TC278810	Unknown
At4g28950	TC265683	Rac GTP binding protein Arac7
At4g29070	TC265576	Unknown
At4g29240	TC272940	Extensin-like protein
At4g32000	TC268678	Serine/threonine protein kinase like protein
At4g32295	TC275317	Unknown
At4g32700	TC257335	Dna-directed dna polymerase family protein
At4g32940	TC268598	Vacuolar processing enzyme gamma
At4g33180	TC267141	Hydrolase, alpha/beta fold family protein
At4g33280	TC266997	Auxin response factor 36
At4g33500	TC273896	BTH-induced protein phosphatase 1
At4g33800	TC275350	Unknown
At4g33820	TC257012	Glycosyl hydrolase family 10 protein
At4g33980	TC264760	Unknown
At4g35335	TC255677	UDP-galactose transporter-like protein
At4g35410	TC263874	Clathrin adaptor complex small chain family protein
At4g35500	TC264619	Protein kinase family protein
At4g36190	TC270120	Serine carboxypeptidase s28
At4g36280	TC275650	Atpase-like domain-containing protein
At4g36630	TC273819	Unknown
At4g36700	TC251621	Globulin-like protein
At4g36910	TC264005	Cbs domain-containing protein
At4g36930	TC255215	Spatula
At4g37050	TC263764	Patatin
At4g38560	TC265789	Phospholipase like protein
At4g38570	TC268772	CDP-diacylglycerol--inositol 3-phosphatidyltransferase
At4g39390	TC263521	Glucose-6-phosphate/phosphate-translocator precursor
At4g39510	TC264187	Cytochrome P450-like protein
At4g39930	TC275293	Unknown
At5g01470	TC266425	Unknown
At5g01670	TC258251	Aldose reductase-like protein
At5g01780	TC276465	Oxidoreductase
At5g03800	TC274332	Limonene cyclase
At5g03860	TC271861	Malate synthase
At5g04010	TC277811	Unknown
At5g04370	TC259603	S-adenosyl-L-methionine:carboxyl methyltransferase family protein
At5g04620	TC265939	Aminotransferase class i and ii family protein
At5g05070	TC276648	Zinc finger (dhhc type) family protein
At5g06830	TC257469	CDK5RAP3-like protein
At5g07190	TC271632	Embryo-specific protein 3
At5g07260	TC255657	Homeobox protein-related
At5g07280	TC276451	Leucine-rich repeat protein kinase
At5g07500	TC256536	Zinc finger transcription factor
At5g07890	TC257058	Myosin heavy chain-related
At5g07990	TC271200	Flavonoid 3'-monooxygenase
At5g08170	TC272770	Peptidyl-arginine deiminase-like protein
At5g08430	TC276564	Swib complex baf60b domain-containing protein
At5g08460	TC266506	GDSL-motif lipase/acylhydrolase-like protein
At5g08480	TC276819	Vq motif-containing protein
At5g08535	TC278158	D111/G-patch domain-containing protein
At5g09640	TC253332	Serine carboxypeptidase
At5g09840	TC277416	Unknown
At5g10460	TC255016	Haloacid dehalogenase-like hydrolase family protein
At5g10700	TC267157	Unknown
At5g11110	TC264890	Sucrose-phosphate synthase
At5g11170	TC268407	DEAD/DEAH box helicase, putative (RH15)
At5g11240	TC254871	Transducin
At5g11310	TC266147	Pentatricopeptide (PPR) repeat-containing protein-like
At5g11320	TC276472	Flavin-containing monooxygenase family protein
At5g11650	TC265511	Hydrolase, alpha/beta fold family protein
At5g11840	TC255973	Unknown
At5g13480	TC278489	Wd-40 repeat family protein
At5g13520	TC263893	Leukotriene-A4 hydrolase-like protein
At5g13570	TC275182	Mrna-decapping enzyme
At5g13690	TC254617	Alpha-N-acetylglucosaminidase
At5g13770	TC267444	Pentatricopeptide (ppr) repeat-containing protein
At5g13790	TC276123	Agamous-like MADS box protein AGL15
At5g15020	TC276581	Paired amphipathic helix repeat-containing protein
At5g15440	TC255045	Circadian clock coupling factor-related
At5g15470	TC263803	Glycosyl transferase family 8 protein
At5g15710	TC274215	F-box family protein
At5g15940	TC268006	Short-chain dehydrogenase/reductase
At5g16070	TC252747	Chaperonin

At5g16310	TC254090	Ubiquitin carboxyl-terminal hydrolase family 1 protein
At5g17040	TC266142	UDP glucose:flavonoid 3-o-glucosyltransferase
At5g18390	TC269516	Pentatricopeptide (ppr) repeat-containing protein
At5g18420	TC272968	Unknown
At5g18840	TC265257	Sugar transporter-like protein
At5g19730	TC276026	Pectinesterase family protein
At5g19850	TC275884	Hydrolase, alpha/beta fold family protein
At5g20040	TC262325	Trna isopentenyltransferase 9
At5g20420	TC268871	Snf2 domain-containing protein
At5g22030	TC273376	Ubiquitin-specific protease-like protein
At5g22470	TC254203	Poly (ADP-ribose) polymerase family protein
At5g22500	TC261409	Acyl coa reductase
At5g22730	TC256230	F-box family protein
At5g22810	TC257381	GDSL-motif lipase/hydrolase-like protein
At5g23520	TC273156	Unknown
At5g24470	TC254105	Pseudo-response regulator 5
At5g24950	TC268102	Cytochrome P450 71A15
At5g26120	TC273481	Glycosyl hydrolase family protein 51
At5g26240	TC265629	Chloride channel protein CLC-d
At5g26760	TC266335	Unknown
At5g26850	TC262492	Unknown
At5g27360	TC275742	Sugar-porter family protein 2 (SFP2)
At5g27610	TC257308	Always early 1 protein
At5g27950	TC274870	BY-2 kinesin-like protein 5
At5g28910	TC268406	Unknown
At5g35450	TC264568	Disease resistance protein
At5g35790	TC273100	Plastidic glucose-6-phosphate dehydrogenase
At5g37580	TC267424	Unknown
At5g37590	TC278438	Kinesin light chain-related protein
At5g38110	TC273956	Anti-silencing protein-like
At5g38160	TC255960	Lipid transfer like protein
At5g38830	TC276343	Cysteine-trna ligase
At5g39130	TC273382	Germin-like protein subfamily 1 member 16 precursor
At5g40420	TC251445	Oleosin
At5g41140	TC274181	Unknown
At5g41150	TC274132	Repair endonuclease
At5g41330	TC256624	Potassium channel tetramerisation domain-containing protein
At5g41580	TC256654	Transcription factor-like protein
At5g42320	TC257350	Zinc carboxypeptidase family protein
At5g42670	TC258664	Agene1 domain-containing protein
At5g42800	TC271692	Dihydroflavonol 4-reductase
At5g43020	TC256511	Leucine-rich repeat transmembrane protein kinase
At5g44050	TC277019	Mate efflux family protein
At5g44310	TC254966	Late embryogenesis abundant protein-like
At5g45690	TC273255	Unknown
At5g45760	TC256088	Transducin family protein
At5g45770	TC265475	Leucine-rich repeat family protein
At5g45830	TC267134	Similarity to tumor-related protein
At5g45850	TC258406	Unknown
At5g46400	TC258157	Unknown
At5g46460	TC277932	Pentatricopeptide (ppr) repeat-containing protein
At5g46540	TC268647	Abc transporter family protein
At5g46870	TC257030	Rna recognition motif (rrm)-containing protein
At5g47150	TC277548	Similarity to SET-domain protein
At5g47670	TC256401	Leafy cotyledon 1-like L1L protein
At5g47720	TC271603	Acetoacyl-coa-thiolase
At5g47800	TC274529	Phototropic-responsive nph3 family protein
At5g48100	TC251742	Laccase
At5g48360	TC273570	Formin homology 2 domain-containing protein
At5g48485	TC273176	Lipid transfer protein (ltp) family protein
At5g49190	TC263845	Sucrose synthase
At5g49950	TC264522	Embryogenesis-associated protein-related
At5g50260	TC263997	Cysteine endopeptidase
At5g50480	TC266431	Transcription factor Hap5a-like
At5g50650	TC265297	Protein transport protein SEC12p-like
At5g50700	TC270936	11-beta-hydroxysteroid dehydrogenase-like
At5g50750	TC263858	Reversibly glycosylated polypeptide RGP-4
At5g50770	TC278807	Short-chain dehydrogenase/reductase (SDR) family protein
At5g50790	TC254997	Mtn3-like protein
At5g51500	TC267997	Pectinesterase
At5g51690	TC263304	1-aminocyclopropane-1-carboxylate synthase (ACC synthase)
At5g51810	TC274586	Gibberellin 20-oxidase
At5g51850	TC276893	Unknown
At5g51870	TC268610	Mads-box protein (agl71)
At5g52330	TC258063	Meprin and traf homology domain-containing protein
At5g52860	TC266648	ABC transporter-like protein
At5g53280	TC275202	Unknown
At5g53440	TC275001	Unknown
At5g53750	TC267570	Cbs domain-containing protein
At5g54740	TC261462	Lipid transfer protein (ltp) family protein
At5g55180	TC257483	Beta-1,3-glucanase-like protein
At5g55240	TC254176	Caleosin-related family protein



At5g55410	TC255953	Lipid transfer protein (ltp) family protein
At5g56300	TC276993	S-adenosyl-L-methionine:salicylic acid carboxyl methyltransferase-like protein
At5g56370	TC266189	F-box family protein
At5g56700	TC278829	F-box family protein
At5g57140	TC264136	Calcineurin-like phosphoesterase family protein
At5g57260	TC275691	Cytochrome P450
At5g57390	TC274827	AP2/EREBP transcription factor
At5g57700	TC255398	BNR/Asp-box repeat family protein
At5g57790	TC268409	Unknown
At5g58080	TC277362	Two-component responsive regulator family protein
At5g59170	TC272223	Cell wall protein precursor, extensin
At5g59190	TC277337	Subtilase family protein
At5g59300	TC266325	Ubiquitin-conjugating enzyme 7 (ubc7), e2
At5g59350	TC273984	Unknown
At5g59590	TC274208	Udp-glucuronosyl/udp-glucosyl transferase family protein
At5g59845	TC264606	Gibberellin-regulated family protein
At5g60610	TC276067	F-box family protein
At5g60760	TC255178	2-phosphoglycerate kinase-related
At5g61150	TC263854	Vernalization independence 4
At5g61390	TC257077	Exonuclease-like protein
At5g62170	TC274808	Unknown
At5g62490	TC254087	Athva22b
At5g62800	TC268787	Seven in absentia (SINA) family protein
At5g62840	TC265463	Phosphoglycerate/bisphosphoglycerate mutase family protein
At5g63000	TC274778	Unknown
At5g63080	TC266876	Transcription factor jumonji (jmc) domain-containing protein
At5g63120	TC254559	ATP-dependent RNA helicase-like protein
At5g63160	TC277582	Speckle-type POZ protein-related
At5g63610	TC263695	Cyclin-dependent kinase cdc2mse
At5g63760	TC254738	ARIADNE-like protein AR115
At5g64200	TC264554	Arginine/serine-rich splicing factor SC35
At5g64900	TC264263	Unknown
At5g65165	TC277079	Succinate dehydrogenase, iron-sulphur subunit, mitochondrial
At5g66180	TC265603	Similar to nol1/nop2/sun family protein
At5g67240	TC255283	Exonuclease
AtMg00180	TC278282	Ccb452 cytochrome c biogenesis orf452
AtMg00270	TC259657	NADH-ubiquinone oxidoreductase chain 6
AtMg00520	TC267036	Maturase
Chloroplast genome	TC262117	Photosystem I assembly protein Ycf3
Chloroplast genome	TC258979	Unknown
AA651576	TC260948	<i>Arabidopsis thaliana</i> 18S rna gene
AC004557	TC253861	Unknown
AC006954	TC269089	Retroelement
AC007534	TC263707	Retroelement
AF074021	TC258129	Transposon
AY072203	TC274611	Transposon
AY088691	TC258748	Unknown
BE520573	TC259174	Unknown
BE520635	TC258019	12S seed storage protein
BE520744	TC268594	Leupaxin
BE520891	TC278755	Unknown
BE520985	TC259388	Unknown
BE521055	TC257247	Unknown
BE521057	TC259044	Unknown
BE521384	TC268730	Unknown
BE521533	TC268737	Unknown
BE521633	TC278198	Unknown
BE522534	TC278435	Unknown
BE523353	TC259401	Unknown
BE524156	TC278329	Unknown
BE525582	TC258944	Unknown
BE523013	TC259196	Unknown
BE523713	TC278028	Unknown
BE528808	TC268646	Unknown
BE530021	TC277499	Unknown
BE530849	TC268861	Unknown

## **APÉNDICE III**

### **Ankyrin repeat-containing proteins in Arabidopsis: characterization of a novel and abundant group of genes coding ankyrin-transmembrane proteins**

Cristian Becerra, Torben Jahrmann, Pere Puigdomènech y Carlos M. Vicent  
Gene 340 (2004): 111 – 121.

# Ankyrin repeat-containing proteins in *Arabidopsis*: characterization of a novel and abundant group of genes coding ankyrin-transmembrane proteins

Cristian Becerra<sup>a</sup>, Torben Jahrmann<sup>a,b</sup>, Pere Puigdomènech<sup>a</sup>, Carlos M. Vicent<sup>a,b,\*</sup>

<sup>a</sup>Departament de Genètica Molecular, IBMB-CSIC, Jordi Girona 18-26, Barcelona 08034, Spain

<sup>b</sup>Departament de Genètica Vegetal, Centre de Cabrils, IRTA, Ctra. de Cabrils s/n, 08348 Cabrils, Barcelona, Spain

Received 3 November 2003; received in revised form 15 May 2004; accepted 1 June 2004

Available online 25 July 2004

Received by W. Martin

## Abstract

Ankyrin repeats are present in a great variety of proteins of eukaryotes, prokaryotes and some viruses and they function as protein–protein interaction domains. We have search for all the ankyrin repeats present in *Arabidopsis* proteins and determined their consensus sequence. We identified a total of 509 ankyrin repeats present in 105 proteins. Ankyrin repeat containing proteins can be classified in 16 groups of structurally similar proteins. The most abundant group contains proteins with ankyrin repeats and transmembrane domains (AtANKTM). Sequence similarity analysis indicates that these proteins are divided in six families. Some of the *AtAnkTm* genes are organized in tandem arrays and others are present in duplicated parts of the *Arabidopsis* genome. The expression of several *AtAnkTm* genes was analyzed resulting in a wide variety of expression patterns even within the same family. The likely functions of these proteins are discussed in comparison with the known functions of proteins with similar organization in other species.

© 2004 Elsevier B.V. All rights reserved.

**Keywords:** Ankyrin; *Arabidopsis*; Protein domain; Transmembrane

## 1. Introduction

Ankyrin repeats (ANK repeats) is a commonly occurring protein repeat present in prokaryotes, eukaryotes and some viruses (Sedgwick and Smerdon, 1999). The primary structure of ANK repeats consists in 33 residues repeated in tandem that built a specific secondary and tertiary structure. Only few of the amino acids are invariant and correspond to hydrophobic positions which are necessary to maintain the secondary structure (Rhode and Bork, 1993; Bork, 1993; Mosavi et al., 2002). ANK repeat tandem arrays consists of two or more repeats separated by less than 20 amino acids (Sedgwick and Smerdon, 1999). ANK repeats are present in proteins involved in very different functions including cell

cycle regulation, mitochondrial enzymes, cytoskeleton interactions, signal transduction or toxins (Sedgwick and Smerdon, 1999).

ANK repeats mediate protein–protein interactions. This function has been experimentally demonstrated in both binding heterologous proteins and mediating homodimerization (Bork, 1993; Lin et al., 1999). For example, ANK repeats are involved in binding together subunits of the GABA-binding protein  $\beta$  (GABP $\beta$ ); the protein I- $\kappa$ B $\alpha$  is almost entirely comprised of ANK repeats and is able to bind the 65 kDa subunit of NF- $\kappa$ B; the  $\alpha$ -latrotoxin from black widow spider venom, which contains 19 ANK repeats, associates with an extracellular protein target; and Su(H) and Deltex proteins bind to Notch ANK repeats.

The folding of the ANK repeats plays an important role for its function (Mosavi et al., 2002). The arrays of ANK repeats consist of pairs on antiparallel  $\alpha$ -helices stacked side by side and linked by a series of intervening  $\beta$ -hairpin motifs. The structure is stabilized by extended antiparallel  $\beta$ -sheets formed between the repeats and by hydrophobic

Abbreviations: ANK, ankyrin.

\* Corresponding author. Departament de Genètica Molecular, IBMB-CSIC, Jordi Girona 18-26, Barcelona 08034, Spain. Tel.: +34-93-4006100; fax: +34-93-2045904.

E-mail address: [cvsmp@cid.csic.es](mailto:cvsmp@cid.csic.es) (C.M. Vicent).

bonds within the repeat and between the neighbouring repeats. The extended  $\beta$ -sheet projects away from the helical pairs almost at right angles, resulting in a L-shaped cross-section. The ability of ANK repeats to bind target proteins involves contact through the tips of the  $\beta$ -hairpins, which are exposed to the solvent, and the surface of the helical bundle facing the ankyrin groove. In general, the residues in the tips of the  $\beta$ -hairpins are not conserved highly in the ANK consensus. As such, they are not structurally constrained and are ideally located to perform binding roles and determine protein interaction specificities.

In plants, few proteins containing ANK repeats have been characterized and the molecular functions of the repeats have not been demonstrated. Here, a comprehensive analysis of the *Arabidopsis* genome detected several genes coding for proteins with ANK repeats that can be classified on 16 groups. We paid special attention to a group of genes coding proteins with ANK repeats and transmembrane domains.

## 2. Materials and methods

### 2.1. Plant material

*Arabidopsis* Col-0 plants were grown in soil, in 22 °C growth chambers, with 18 h light days. Plants used for RNA extractions from roots were grown on 0.8% (w/v) Murashige and Skoog basal salt mixture agar plates in 22 °C growth chambers under 18 h light days.

### 2.2. Sequence acquisition and analysis

The initial set of ANK repeats containing proteins was compiled from SMART v3.5 (<http://smart.embl-heidelberg.de/>) and TAIR Protein Search <http://www.arabidopsis.org>). The phasing of the repeats was that proposed by Michaely and Bennett (1992). Additional repeats were obtained using REP v1.1 (<http://www.embl-heidelberg.de/~andrade/papers/rep/search.html>). A multiple alignment of ANK repeats was obtained using CLUSTALW in order to construct representative sequences. With these sequences we screened protein databases using BLAST (<http://www.ncbi.nlm.nih.gov/BLAST/>) and identified unique hits, removing duplications from our data set caused by the multiple identification numbers frequently assigned to the same DNA or protein sequence in the databases. The search was conducted in three steps, in each one searching and adding new repeats and recalculating consensus patterns. ANK repeats do not have high sequence conservation (Bork, 1993) and using this system we ensured isolation of all putative ANK repeats. However, we also risk including false hits. In order to remove noisy hits we took into account that ANK repeats are always found in arrays of at least two repeats separated by not more than twenty amino acids. We removed all the “isolated” ANK repeats obtained. We also removed all partial repeats and those putative ANK repeats that do not conserve at least two of the six most conserved amino acids defined by Bork (1993). In a final round, several candidates with weak signals were examined in detail. The less conserved ANK repeats were added only when they occur

Table 1  
PCR primers used for RT-PCRs

Atg Number	Primer sequence 5'	Primer sequence 3'
At4g03440	CTTGGATTTGCTACGTCGTAGCC	GAGATACTGCTCTCCACTCAGCC
At4g03450	CTTGACACGCGGCTCTAAAGGC	ACCTCTTGGCGAACAGTGCAC
At4g03480	CAGATTCCGCTTCATGTGGCCGC	CAGGCAGCTGTCTCCATTTGGCG
At4g05040	GCAGGTAACAATGACCTTGAAGGG	CCTGCAGCAAACGTCATTGTGGC
At4g14400	GACAACGTGGACCGTGAAGTGAGG	AGAGCCGCTACCACGAGAAGAGC
At1g14480	CGGGATGGATCCAGAGAATGAGCC	GGAGGGCCATCTGATAAGTGGCTG
At1g14500	GCCAAAAAGATTCTGCTTCCACCG	GCTGGAGTGCAGTTTGATAAGTGG
At4g10720	GCGAGAAAACCTAACACATACGGG	GGTTATGTAAGATATCTAGGGCGG
At4g11000	GCTAAACGTGTCAGTTTCAGCCC	GGAAGGTGAGCACTGAGAGATAGC
At5g15500	AACGCCGACGGACTTACAC	ATTCCAAAAACAAAACCTACC
At5g51160	GGGGTTGAAAAGAAGCTTTGCCG	GCTTCGCTCGGAAACATAACCAGC
At5g54610	GGGTGGATGCAGAAAATGCGCG	GGGTGCGAGTCTGAAAAGTGGC
At5g54620	GCGAGACTGCTCTACATATTGCGG	GCGATCTCAAGAAGACAGATGCGG
At2g24600	CTTGAGCTTGTGAGGGAGAAGG	CTCTCTCCGTAACGCGTACGG
At5g54710	GGCCCAGAGTGCAAACATACGCC	CACCACCAGGAGGGTTTATCCCG
At1g07710	GGGAAACAGAAACAGTCAAGCG	CCATGTGAAGGGCTGTTTGGCCC
At2g01680	CGCTTTTCATGTCGCTGCCAAGCG	CCGTAAAGGTCAGCAATAAGCTCG
At3g09550	GTTGCTCCTTCGAGCTGATCCGG	AGCTCCTTGGCGATGCCATCGACG
At3g12360	GCTGAAGTTGCGGAGATTCGAGC	GTCAAGCGCTGTCTTGTGATCGC
At5g02620	CCGGAACCAAAGCCAAGAACGGC	AGCTGAGTATGCACCTCATGGCC
At5g60070	AGGGTCAGACGCCATTCACATG	CGTTTCTCCGAGAGATTGTCCCG
At5g04690	AACACGCAGATGGAGTGGCTCG	AACGTTCTTCAAGCCCAAGGCC
At5g09810 (Actin)	GGCCGATGGTGAGGATATTC	CTGACTCATCTACTACTCTC
At2g40170 ( <i>AtEm6</i> )	GGCGTCTCAACAAGAGAAGAAGC	GGGGAAGTTTGATTTAGGTCTTG

between clearly identified repeats or were located in the extreme of a tandem array.

The analysis of the presence of additional domains other than ANK repeats was performed using SMART v3.5 (<http://smart.embl-heidelberg.de/>). The alignment of the protein sequences was done with CLUSTALW and phylogenetic analysis were performed using the neighbor-joining method. The analysis of chromosomal duplications in the *Arabidopsis* genome was done using the on-line facility provided by Ken Wolfe (<http://wolfe.gen.tcd.ie/athal/dup>).

The position of the introns, start and stop codons of the predicted AtANKTM proteins was checked by comparing the genomic sequences with the cDNA sequences deposited into the GenBank, comparing the proteins of the same family in the alignments and by direct sequencing of RT-PCR fragments. We found small differences from the prediction in sixteen of the forty genes. All the analyses shown here were done with the reviewed sequences.

### 2.3. RNA extraction and RT-PCR

Total RNAs were extracted from frozen organs of *Arabidopsis* as described (Vicent and Delseny, 1999) and treated with DNase I (RNase-free DNaseI, Promega). Total pretreated RNA (2 µg) was reverse transcribed with Omniscript reverse transcriptase kit (Qiagen) using an oligo-dT primer. cDNAs were amplified with specific primers designed flanking introns (Table 1). Reaction controls with nonreverse transcribed RNA were also used to detect gDNA contamination. The actin gene was used as a control of RNA loading. *AtEm6* gene was used as a control of expression in mature seeds (Vicent et al., 2000). PCR reactions were performed using 0.2 mM each dNTP, 360 µg/ml BSA and 1 pmol µl<sup>-1</sup> each primer in a final volume of 50 µl. The reaction mixtures were heated to 95 °C for 5 min, followed by 28 cycles of 94 °C for 30 s, 55 °C for 30 s, and 72 °C for 90 s. Reactions were completed with one incubation at 72 °C for 10 min. Reactions were performed in a Minicycler (MJ Research, Waltham, MA) thermal cyclers. Reaction products were cloned in pGEM-T (Promega) for sequencing.

## 3. Results

### 3.1. Generation of a nonredundant ankyrin containing protein set

The target of this work is to identify all the ankyrin repeats in *Arabidopsis* proteins using database searches. Our searching scheme (see Section 2) allowed us to identify a total of 509 ANK repeats coded by 105 genes. The number of ANK repeats in the same array ranks between 2 and 10 and the average is 4.5. Some of the proteins contain two separate arrays of ANK repeats. A summary, including Atg number for each of these proteins is provided in Table 2. For

Table 2  
*Arabidopsis* ankyrin repeat containing proteins

EN	Atg no.	Fa. <sup>a</sup>	Name	Reference
<i>Cluster A: Ankyrin-transmembrane proteins</i>				
1	At1g03670	1	n.a.	n.a.
2	At4g03440	1	n.a.	n.a.
3	At4g03450	1	n.a.	n.a.
4	At4g03460	1	n.a.	n.a.
5	At4g03470	1	n.a.	n.a.
6	At4g03480	1	n.a.	n.a.
7	At4g03490	1	n.a.	n.a.
8	At4g03500	1	n.a.	n.a.
9	At4g05040	1	n.a.	n.a.
10	At4g14390	1	n.a.	n.a.
11	At4g14400	1	ACD6	Lu et al., 2003
12	At1g14480	2	n.a.	n.a.
13	At1g14500	2	n.a.	n.a.
14	At4g10720	2	n.a.	n.a.
15	At4g11000	2	n.a.	n.a.
16	At5g15500	2	n.a.	n.a.
17	At5g51160	2	n.a.	n.a.
18	At5g54610	2	n.a.	n.a.
19	At5g54620	2	n.a.	n.a.
20	At1g10340	3	n.a.	n.a.
21	At1g34050	3	n.a.	n.a.
22	At2g24600	3	n.a.	n.a.
23	At5g50140	3	n.a.	n.a.
24	At5g54700	3	n.a.	n.a.
25	At5g54710	3	n.a.	n.a.
26	At5g54720	3 <sup>b</sup>	n.a.	n.a.
27	At1g05640	4	n.a.	n.a.
28	At1g07710	4	n.a.	n.a.
29	At2g01680	4	n.a.	n.a.
30	At2g31820	4	n.a.	n.a.
31	At3g09550	4	n.a.	n.a.
32	At3g12360	4	n.a.	n.a.
33	At5g02620	4	n.a.	n.a.
34	At5g60070	4	n.a.	n.a.
35	At3g18670	5	n.a.	n.a.
36	At3g54070	5	n.a.	n.a.
37	At5g04690	5	n.a.	n.a.
38	At5g35830	5 <sup>b</sup>	n.a.	n.a.
39	At2g14250	6 <sup>b</sup>	n.a.	n.a.
40	At5g20350	6	n.a.	n.a.
<i>Cluster B: Proteins with only ankyrin repeats</i>				
41	At1g04780	1	n.a.	n.a.
42	At1g11740	1	n.a.	n.a.
43	At1g62050	1	n.a.	n.a.
44	At3g04470	1	n.a.	n.a.
45	At3g24210	1	n.a.	n.a.
46	At2g17390	2	ART2	Peck et al., 2001
47	At4g35450	2	ARP2	Yan et al., 2002
48	At5g40160	3	EMB506	Albert et al., 1999
49	At5g66055	3	AKR	Zhang et al., 1992
50	At5g07840	4	n.a.	n.a.
51	At5g61230	4	n.a.	n.a.
52	At3g01750	5	n.a.	n.a.
53	At3g04140	5	n.a.	n.a.
54	At5g65860	6	n.a.	n.a.
55	At4g19150	7	n.a.	n.a.
56	At2g03430	8	n.a.	n.a.
57	At5g12320	9	n.a.	n.a.
58	At3g09890	10	n.a.	n.a.

(continued on next page)

Table 2 (continued)

EN	Atg no.	Fa. <sup>a</sup>	Name	Reference
<i>Cluster C: Proteins with BTB domain</i>				
59	At1g64280	1	NPR1	Cao et al., 1997
60	At2g41370	1	NPR1	Cao et al., 1997
61	At3g57130	1	NPR1	Cao et al., 1997
62	At4g19660	1	NPR1	Cao et al., 1997
63	At4g26120	1	NPR1	Cao et al., 1997
64	At5g45110	1	NPR1	Cao et al., 1997
65	At2g04740	2	n.a.	n.a.
<i>Cluster D: Protein kinases</i>				
66	At1g14000	1	APK-like	Chinchilla et al., 2003
67	At2g31800	1	Atapk2	Chinchilla et al., 2003
68	At3g58760	1	APK-like	Chinchilla et al., 2003
69	At3g59830	1	Atapk3	Chinchilla et al., 2003
70	At4g18950	1	APK-like	Chinchilla et al., 2003
71	At2g43850	2	Atapk1	Chinchilla et al., 2003
72	At5g13530	3	n.a.	n.a.
<i>Cluster E: Zinc-finger proteins</i>				
73	At2g40140	1	n.a.	n.a.
74	At2g41900	1	n.a.	n.a.
75	At3g55980	1	n.a.	n.a.
76	At5g12850	1	n.a.	n.a.
77	At5g58620	1	n.a.	n.a.
78	At3g28880	2	n.a.	n.a.
<i>Cluster F: Potassium channels</i>				
79	At2g25600	1	AKT	Pilot et al., 2003
80	At2g26650	1	AKT	Pilot et al., 2003
81	At3g02850	1	AKT	Pilot et al., 2003
82	At4g22200	1	AKT	Pilot et al., 2003
83	At4g32500	1	AKT	Pilot et al., 2003
84	At5g37500	1	AKT	Pilot et al., 2003
<i>Cluster G: Ring Finger proteins</i>				
85	At3g23280	1	n.a.	n.a.
86	At4g14365	1	n.a.	n.a.
87	At5g07270	2	n.a.	n.a.
88	At5g57740	2	n.a.	n.a.
89	At2g28840	3	n.a.	n.a.
<i>Cluster H: ARF GTPase-activating domain-containing protein</i>				
90	At1g10870	1	n.a.	n.a.
91	At1g60860	1	n.a.	n.a.
92	At5g13300	1	n.a.	n.a.
93	At5g61980	1	n.a.	n.a.
<i>Cluster I: Calmodulin binding motif-containing protein</i>				
94	At1g67310	1	CAMTA	Bouché et al., 2002
95	At2g22300	1	CAMTA	Bouché et al., 2002
96	At5g09410	1	CAMTA	Bouché et al., 2002
97	At5g64220	1	CAMTA	Bouché et al., 2002
<i>Cluster J: Acyl-CoA binding protein</i>				
98	At4g27780	1	ACBT	Chye et al., 2000
99	At5g53470	1	ACBT	Chye et al., 2000
<i>Cluster K: Chromodomain protein</i>				
100	At2g47450	1	CAO	Klimyuk et al., 1999
<i>Cluster L: Helicase</i>				
101	At1g06670	1	DEAH	Isono et al., 1999

Table 2 (continued)

EN	Atg no.	Fa. <sup>a</sup>	Name	Reference
<i>Cluster M: Protein with RCC-1 domains</i>				
102	At3g03790	1	n.a.	n.a.
<i>Cluster N: Protein with tetratricopeptide repeats</i>				
103	At3g04710	1	n.a.	n.a.
<i>Cluster O: Protein with pH domain</i>				
104	At5g14230	1	n.a.	n.a.
<i>Cluster P: Protein with ATPase associated motif</i>				
105	At3g24530	1	n.a.	n.a.

<sup>a</sup> Families inside each cluster based on sequence similarity.

<sup>b</sup> Proteins that only have ankyrin repeats but by sequence similarity they are grouped with proteins containing additional domains.

convenience, we assigned an “entry number” (EN) for each protein.

### 3.2. Determination of a consensus sequence of *arabidopsis* ankyrin repeats

A consensus sequence was determined based on the alignment of the sequences of the 509 *Arabidopsis* ANK repeats (Table 3). None of the positions was completely conserved but we identified 7 residues with at least 50% conservation across all repeats. Five of them are located between positions 2 and 9, which is the region of overall higher conservation. All these residues have been previously reported as highly conserved based largely on animal proteins and the presence of strictly conserved hydrophobic and hydrophilic positions has also been observed (Michaely and Bennett, 1992; Bork, 1993).

### 3.3. Phylogenetic and structural analysis of ankyrin-containing proteins in *Arabidopsis*

The 105 ANK repeat containing proteins were analyzed for the presence of additional recognizable protein motifs and 16 clusters of structurally related proteins were identified (Table 4). Protein sequences of each cluster were aligned and a similarity tree was generated (not shown), giving rise to a classification in families of sequence related proteins. Thirty seven *Arabidopsis* genes code for proteins containing ANK repeats and transmembrane domains (cluster A) whereas 21 code proteins with only ANK repeats as recognizable motif (cluster B). Genes EN26, EN38 and EN39 code for proteins structurally belonging to cluster B, however, their sequences are more related to some proteins in cluster A. They probably correspond to truncated forms of proteins of the cluster A and we include them in this group.

Among the 105 proteins isolated only 31 have been previously characterized to any extent. Clusters with higher number of genes are those with a lower proportion of studied genes. Only one of the 40 genes in cluster A has



Table 3  
Consensus sequence of *Arabidopsis* Ankyrin repeats

aa <sup>a</sup>	<i>Arabidopsis</i>		Consensus sequence by	
	Consensus	Most abundant amino acids	Bork, 1993	Michaely and Bennett, 1992
1	Hyphyl.	D 21%	–	–
2	G	G 63%	Turn-like or polar	G
3	Hyphyl.	–	–	–
4	T	T 56%	T or S	T
5	P, A	P 37%; A 29%	P	A, P
6	L	L 77%	L	L
7	H	H 64%	H	H
8	Hyphob.	L 23%	Hydrophobic	L, I, V
9	A	A 84%	A	A
10	A	A 48%; V 20%	Hydrophobic	A, S
11	Hyphyl.	–	–	R, Q, K
12	Hyphyl.	–	–	–
13	G	G 62%	Turn-like or polar	G, N
14	Hyphyl.	H 24%	Turn-like or polar	H, N
15	–	–	–	V, L, T
16	Hyphyl.	E 29%	Turn-like or polar	E, D
17	Hyphob.	V 23%; I 20%	hydrophobic	V, I, M
18	V	V 45%	hydrophobic	V, A
19	Hyphyl.	K 27%	Turn-like or polar	K, E, R
20	–	L 18%	–	L, V
21	L	L 63%	L	L
22	L	L 46%	L	L
23	Hyphyl.	E 19%	Turn-like or polar	D, K, Q, E
24	Hyphyl.	–	–	–
25	–	G 30%	Turn-like or polar	G
26	–	A 24%; P 18%	–	A
27	Hyphyl.	D 20%	–	D, N, S
28	–	L 19%	–	V, P, I
29	–	–	–	N, D
30	–	–	–	A
31	–	–	–	–
32	Hyphyl.	D 19%	–	T, D, N
33	Hyphyl.	N 19%	–	K

The consensus contains a single amino acid when it represents alone more than 40% and amino acid class when it represents more than 60%. Hyphen means no special amino acid or amino acid type in that position. Most abundant amino acids are only indicated if they represent 18% or more of the total. Percentages refer to the 509 ankyrin repeats identified in this study.

<sup>a</sup> Amino acid position.

been recently characterized (Lu et al., 2003). For this reason we decided to study this family in a greater depth in Section 3.4.

### 3.3.1. Proteins with only ankyrin repeats

Twenty one *Arabidopsis* genes encode for proteins with ANK repeats as the only recognizable motif (Table 2): families B1 to B10 and genes EN26, EN38 and EN39. The size of the encoded proteins range from 144 to 664 amino acids and the number of ANK repeats from 2 to 10. In the most extreme case, ANK repeats comprise 87% of the

protein (EN57), in other cases ANK repeat arrays are concentrated in the N- or C-terminal regions. Four of these genes have been previously characterized. Two of them (EN46 and EN47) code proteins similar to tobacco ANK1 and are involved in pathogen defense (Peck et al., 2001; Yan et al., 2002) and the roles of EN48 and EN49 are related to embryogenesis and development (Zhang et al., 1992; Albert et al., 1999).

### 3.3.2. BTB domain cluster

Analysis of the *Arabidopsis* genome revealed seven genes that encode proteins with a BTB domain and ANK repeats. The BTB domain, also known as POZ (poxvirus and zinc finger), is known to be a protein–protein interaction motif found at the N-terminus of several C2H2-type transcription factors as well as Shaw-type potassium channels (Bardwell and Treisman, 1994). They are divided into two families (C1 and C2) with a different order in the domains. The C1 family contains six proteins in which the BTB domain is located in the N-terminus and the ANK repeats in the C-terminus. These genes encode proteins similar to NPR1, a protein involved in the control of the onset of systemic acquired resistance to a broad spectrum of pathogens (Cao et al., 1997). Proteins similar to NPR1 have been observed in other plant species such as rice, *Brassica* and tobacco, but any protein with a similar domain structure has been found in animals or fungi. The C2 family contains one gene coding for a protein with the domains in the reverse order to the C1 proteins. Proteins with similar domain organization have been observed in animals and fungi, but their functions are unknown except for a human elongation factor 1A binding protein (Unoki and Nakamura, 2001).

### 3.3.3. Protein kinases

Seven genes code proteins with ankyrin repeats and protein kinase domains, divided into three families. In D1 (five genes) and D2 (one gene) families, the ANK repeats are located in the N-terminus and the kinase domain in the C-terminus. D1 proteins are similar to the *Medicago* APK ankyrin protein kinase (Chinchilla et al., 2003). Structural homologues of D1 and D2 proteins also exist in animals, for example, the human cardiac ankyrin repeat kinase (Acc. NM\_015978). The D3 family contains a single gene (EN72) that codes a protein with an N-terminal RING finger domain, a kinase domain in the middle and C-terminal ANK repeats. The gene EN66 seems to encode a tyrosine kinase, but the kinase specificity of the other proteins is not known.

### 3.3.4. Zinc-finger proteins

Six genes code proteins with ANK repeats and zinc-finger domains. They are divided into two families (E1 and E2). Family E1 contains five genes coding proteins with short arrays of two or three ANK repeats at the N-terminus and one or two zinc-fingers in the central part of the protein.

Table 4  
Clusters of ANK containing Proteins in *Arabidopsis*

Class	Description	Number of genes	Scheme of a representative protein
A	Proteins with transmembrane domains	40	—[A][A][A][A][A]—[T][T][T][T]—
B	Proteins only with ankyrin repeats	18	—[A][A][A][A][A]—
C	Proteins with ankyrin and BTB domain	7	—[BTB]—[A][A]—
D	Kinases	7	—[A][A][A]—[KIN]—
E	Proteins with Zinc finger	6	—[A][A]—[Z][Z]—
F	Potassium channel proteins	6	—[T]—[IT]—[cNMP]—[A][A][A][A]—
G	Proteins with Ring Finger	5	—[A][A][A][A][A]—[R]—
H	Proteins with ARF GTPase-activating domain	4	—[BAR]—[PH]—[ARF]—[A][A]—
I	Proteins with calmodulin binding motif	4	—[CG-1]—[A][A]—[Iq][Iq]—
J	Proteins with Acyl CoA binding domain	2	—[T]—[ACBP]—[A][A]—
K	Proteins with chromodomain	1	—[C][A][A][A]—[C]—[C]—
L	Helicase	1	—[R]—[DEXD]—[A][A]—[HE]—[HA]—
M	Regulator of chromosome condensation	1	—[A][A]—[R]—[R]—[R]—[R]—
N	Protein containing Tetratricopeptide repeats	1	—[A][A][A][A][A][A]—[Tt][Tt][Tt]—
O	Proteins with PH domain	1	—[A][A][A]—[A]—[A][A]—[PH]—
P	Proteins with ATPase associated domain	1	—[A][A][A][A]—[AAA]—

A ankyrin repeat; T transmembrane; BTB, Broad-Complex, Tramtrack and Bric a brac; KIN, protein kinase domain; IT, ion transport protein domain; Z, zinc-finger domain; cNMP, cyclic nucleotide-monophosphate binding domain; R, RING finger; BAR, BAR domain; PH, Pleckstrin homology domain; ARF, Putative GTP-ase activating proteins for the small GTPase; CG-1, CG-1 domain; Iq, Short calmodulin-binding motif containing conserved Ile and Gln residues; ACBP, Acyl CoA binding protein domain; C, chromodomain; R, putative single-stranded nucleic acids-binding domain; DEXD, DEAD-like helicases superfamily; HE, helicase superfamily c-terminal domain; HA, helicase associated domain; R, regulator of chromosome condensation; Tt, tetratricopeptide repeats; AAA, ATPase associated domain.

Similar proteins are present in rice. The E2 family contains one gene that codes a protein with similar domains but having an array of six ANK repeats. The functions of none of these proteins have been determined.

### 3.3.5. Potassium channels

The first plant protein described containing an ANK repeat was AKT1 (EN80) which codes a protein similar to a shaker-like  $K^+$  channel located in the plasma membrane (Sentenac et al., 1992). Shaker potassium channels play an important role in the uptake of  $K^+$  from the soil. The *Arabidopsis* genome contains nine genes coding Shaker potassium channels and six of them contain ANK repeats (Pilot et al., 2003). Plant Shaker channels share a common structure: a hydrophobic core composed of six transmembrane segments, a long cytoplasmic C terminal region containing a putative cyclic nucleotide binding domain, and a KHA domain. Many channels, but not all, also contain ANK repeats between the putative cyclic binding domain and the KHA domain. Similar proteins have been described in many other plant species including dicots and monocots (Pilot et al., 2003).

### 3.3.6. Ring finger proteins

There are six *Arabidopsis* genes encoding proteins with RING domains and ANK repeats. One of them also has a kinase domain and for this reason has been included in the family D3 (EN72). The remaining five contain four to six ANK repeats at the N-terminus and a RING finger at the C-terminus. They are divided into three families (G1–G3). Similar proteins have been found in rice and in animals, but the functions are unknown.

### 3.3.7. ARF GTPase-activating domain-containing protein

Four *Arabidopsis* genes with a similar complex organization were identified (Family H1), from N to C terminus: a BAR domain, a PH domain, a GTPase activating domain and two to three ANK repeats. None of these genes have been studied in any plant species.

### 3.3.8. Calmodulin binding motif-containing protein

Four *Arabidopsis* genes (family I1) code proteins with a similar organization: an N-terminal CG-1 domain, two or three ANK repeats in the central region, and two calmodulin binding motifs in the C-terminus. They were named CAMTAs (Calmodulin-binding transcription activators) and have been described in *Arabidopsis* and other plant species (Bouché et al., 2002). CG-1 domains are highly conserved with about 130 amino acid residues containing a predicted bipartite NLS and named after a partial cDNA clone isolated from parsley encoding a sequence-specific DNA-binding protein (da Costa e Silva, 1994). CG-1 domains are associated with CAMTA proteins.

### 3.3.9. Acyl-CoA binding protein

Two different genes encoding cytosolic acyl-CoA-binding proteins were identified (family J1). These proteins also contain ANK repeats in the C-terminal region (Chye et al., 2000) and a transmembrane motif at the N-terminus.

### 3.3.10. Chromodomain protein

The *Arabidopsis* CAO gene (chlorophyll a/b binding protein harvesting-organelle specific protein) codes a protein with ANK repeats and chromodomains (Klimyuk et al., 1999). This is a nuclear gene encoding a chloroplast signal



recognition particle, which is part of a protein complex. The ANK repeats are necessary for the formation of the complex.

### 3.3.11. Helicase

Gene EN101 codes a protein similar to the DEAH family of RNA/DNA helicases (Isono et al., 1999). The protein contains two ANK repeats.

### 3.3.12. Other proteins

Four more genes coding proteins with ANK repeats have been found. They also contain some other recognizable protein motifs such as four RCC-1 domains (EN102), three tetratricopeptide repeats (EN103), a PH domain (EN104) and an ATPase associated motif (EN105). None of these, or similar proteins, have been characterized in plants and their functions are unknown.

### 3.4. Ankyrin-transmembrane proteins in *Arabidopsis*. Phylogenetic analysis and genome distribution

Thirty-seven proteins in cluster A contain four to eleven N-terminal ANK repeats and two to five C-terminal transmembrane domains (Table 2). These proteins were named AtANKTM proteins (*Arabidopsis thaliana* ankyrin transmembrane). Three additional genes (EN26, EN38 and EN39) code for proteins without transmembrane domains but with sequence similarity to the ANK repeat region of the AtANKTM proteins.

The predicted amino acid sequences of the ANKTM proteins were aligned and a similarity tree constructed. Sequence comparison demonstrates that they are divided into six distinct families named 1–6. Similar results were obtained when the analysis was performed using only the N-terminal ankyrin containing part of the proteins (Fig. 1A) or only the C-terminal transmembrane containing part. Based on the alignments, a consensus protein model for each family is shown in Fig. 1B. Some of the ANK repeats are present in most or all the proteins of the family but some are lacking in some proteins due to insertions, deletions or single mutations. The position and number of transmembrane domains is conserved in all proteins of the same family except for the three truncated proteins.

We analyzed the intron distribution of all the AtANKTM genes reported here (data not shown). Only one gene has no introns. The great majority of introns (89 of 96) are located in the ankyrin coding regions of the genes. The position of the introns is not correlated with the position of the ANK repeats as has been observed for some mammalian and plant genes (Albert et al., 1999). Forty-one of the introns interrupt the region coding for an ANK repeat. We analyzed whether the intron/exon number and distribution patterns are related to the phylogenetic distribution in families. Although some introns are present in more than one gene of the same family, in general, their position and number are not conserved.

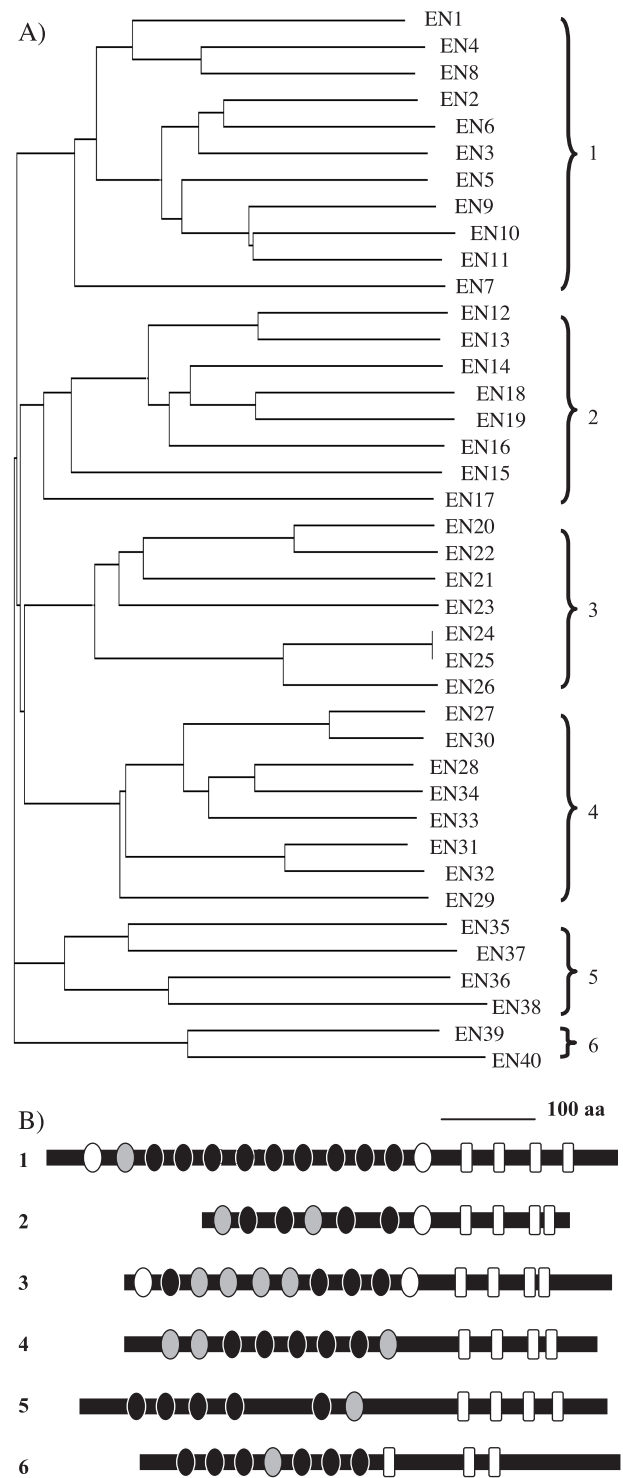


Fig. 1. Families of *Arabidopsis* proteins containing Ankyrin repeats and transmembrane domains (ANKTM). (A) Neighbor-joining tree of the N-terminal region of the ANKTM proteins containing the ankyrin repeats. Numbers in the right indicate the different families. (B) Schematic representation of the hypothetical consensus proteins of each of the six ANKTM protein families. Circles represent ankyrin repeats. Black circles represent ankyrin repeats present in >90% of the proteins of the family, grey circles present in 50–89% of the proteins and white circles present in <50%. Empty rectangles represent transmembrane domains.

*AtAnkTm* genes are distributed in all chromosomes although not uniformly (Fig. 2). Whereas chromosome II and III contain four genes, chromosome V contains 13. There are some areas with a high density of genes such as at the bottom of chromosome V and the top of chromosome I. Conversely, there are large regions that are devoid of *AtAnkTm* genes, including the bottom of chromosomes I and IV. There are five cases of two or more genes arranged in tandem (Fig. 2). Three of them correspond to couples of genes (EN10 and EN11, EN12 and EN13, and EN18 and EN19), one include three genes (EN24, EN25 and EN26) and finally, there are seven genes arranged in tandem in chromosome IV (EN2 to EN8). All these genes are closely related in the phylogenetic analysis (Fig. 1A) with the exception of the group of seven genes which are not so closely related although they belong to the same family.

Analysis of the inter- and intrachromosome duplicated areas of the *Arabidopsis* genome indicates that there are three cases of correlation between gene localization and genome duplications. The gene EN27 in chromosome I seems to be duplicated in chromosome II (EN30). Accordingly, genes EN27 and EN30 are closely related in the phylogenetic analysis (Fig. 1A). Genes EN28, EN33 and EN34 are located in an area repeated three times in the *Arabidopsis* genome, twice in chromosome V and one in chromosome I. These genes are also closely related according to the phylogenetic analysis. Finally, gene EN1 is located in a region of chromosome I duplicated in the area of chromosome IV containing the group of seven *AtAnkTm* genes in tandem. In the phylogenetic analysis EN1 is closely related to EN4 and EN8.

### 3.5. Expression analysis of the *AtAnkTm* genes

Total RNA was extracted from different *Arabidopsis* organs (roots, leaves, flowers, stems, caulinar leaves and siliques at three different stages of development) and ana-

lyzed by semiquantitative RT-PCR using pairs of primers specific to 22 of the *AtAnkTm* genes (Table 1) (Fig. 3). Controls for DNA contamination, RNA integrity and equalization of the quantities of cDNA are described in Section 2. The actin gene was expressed at similar levels in all organs and the PCR amplification using *AtEm6* primers was according to the previous data (a gene specifically expressed in seed maturation) (Vicent et al., 2000). On the other hand, databases were screened for *Arabidopsis* ESTs corresponding to the *AtAnkTm* genes (Table 5). We can not observe any clear correlation between gene family and organ or stress-response transcription specificity.

RT-PCR amplification gave bands for 13 of the genes in at least one of the samples. Primers for genes EN25 and EN32 produced bands in all the organs tested except for mature siliques. EN25 is highly expressed in leaves and stem and an EST was found corresponding to an “above-ground tissues” library. Gene EN32 is also expressed in all organs but is highly expressed in leaves. Primers for gene EN29 produced amplification in all tissues except roots and in a low level in stems. Several ESTs from different tissues were found in databases corresponding to this gene. Other genes from which we detected amplification in several organs are EN12, EN16, EN18, EN19, EN22 and EN33.

Some of the genes seem to have more specific patterns of expression. For example, primers for genes EN14 and EN17 amplified only in roots and primers for gene EN11 only from leaves. The pattern of expression of EN11 is consistent with the expression pattern reported by Lu et al. (2003) for the ACD6 gene.

RT-PCR did not amplify bands using eight pairs of primers (genes EN2, EN3, EN6, EN9, EN13, EN15, EN28, EN31 and EN34). Accordingly, there are not corresponding ESTs for genes EN6, EN13, EN15 and EN34. ESTs corresponding to genes EN2, EN3 and EN9 came from stressed tissues, conditions that we did not tested. Gene EN3 has a corresponding EST from root and

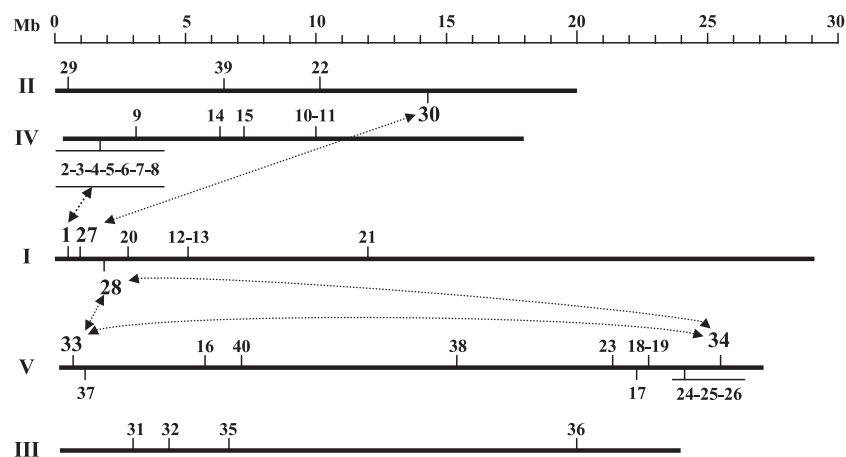


Fig. 2. Chromosomal distribution and duplication events for *Arabidopsis AnkTm* genes. Deduced chromosomal positions of the *AtAnkTm* genes are indicated by EN. The scale is in Megabases (Mb). Numbers separated by hyphens represent genes arranged in tandem. Connecting pointed lines mark a correlation between duplicated genomic region and the presence of *AnkTm* genes.

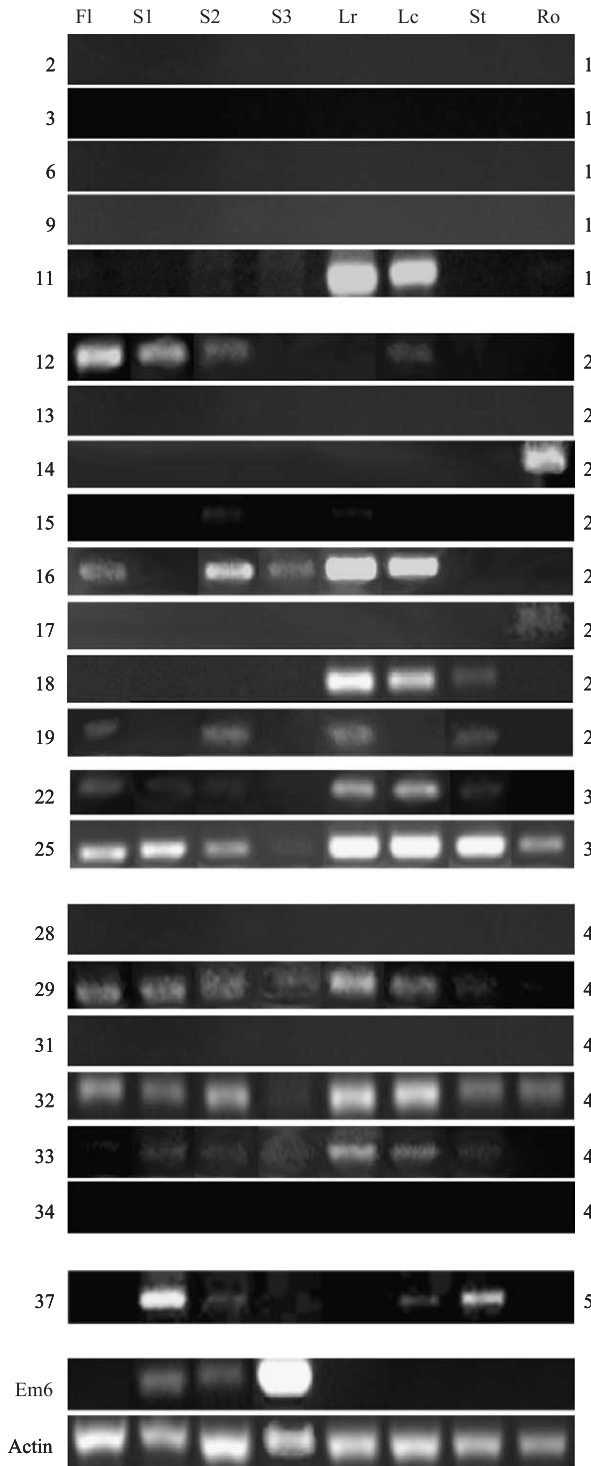


Fig. 3. RT-PCR analysis of the expression profiles of 22 *Arabidopsis* genes encoding ANKTM proteins. Ethidium bromide-stained 1.5% agarose gels showing RT-PCR products. The EN number of the corresponding genes is shown in the left. The corresponding gene family is shown on the right. RNAs used were extracted from flowers (Fl), immature siliques 1–7 dap (S1), intermediate siliques 8–14 dap (S2), mature siliques 14–21 dap (S3), rosette leaves (Lr), caulinar leaves (Lc), stems (St) and roots (Ro). Total RNA (2 µg) was used to synthesize cDNA. A fraction (1/20) of the synthesized cDNA was used to amplify gene transcripts by PCR. The bands shown corresponded to the expected size in each case. *Actin* and *AtEm6* genes were used as controls.

Table 5  
Summary of information on ESTs of the *AtAnkTm* genes

EN	F <sup>a</sup>	ESTs. Accession Number and origin
1	1	None
2	1	AV798477; dehydration and cold
3	1	AI997466; root BE662952; ozone
4	1	None
5	1	None
6	1	None
7	1	None
8	1	None
9	1	AV793119; cold
10	1	None
11	1	AI998090; leaf AV440557; aboveground AI100255, H37081, T42092, T46084; Pooled BE662757; ozone AV822072, AV793013; cold AV794986, AV826452, AV830548; dehydration AV813132, AV815244, AV807835; dehydration and cold AU231325; various stresses
12	2	AV549133, AV537342; root
13	2	None
14	2	None
15	2	None
16	2	AU230590, AU239298; silique and flower BE038658; salt
17	2	AI996466; root
18	2	AU229652; AU238464; dehydration
19	2	None
20	3	AV798019, AV816532, AV827371; dehydration and cold BE662714; ozone. BU635941, CB074318; infected leaf
21	3	None
22	3	AI996591; root. AV825484, AV791380; cold AV795644; dehydration. BE662700, BE662788; ozone
23	3	None
24	3	AU237468, T45361; pooled
25	3	AV522917; aboveground. BE662949; ozone
26	3	AV528599; aboveground BF381542; ozone
27	4	None
28	4	AI996003; inflorescence
29	4	AA728512; inflorescence. AI996553; root AV557218, AV558991, AV561448; green siliques AV439689, AV441879; aboveground AI993199, R64963, T43424, T46195; pooled AV784945, AV787218, AV823873; dehydration and cold
30	4	CB185927; infected leaf
31	4	T04286, AA585945; pooled
32	4	H36055; pooled
33	4	AV534234; flower buds. AV782571, AV821976; cold AV803115, AV815435; dehydration and cold
34	4	None
35	5	None
36	5	None
37	5	AV564960; green siliques
38	5	None
39	6	N96603; pooled
40	6	R90543, AA605447, AI994371; pooled AV519957; aboveground BE521378, BE521377, AV808021, AV814010; mature seed AU227998, AU236986; silique and flower AV566955; green siliques. AV539138; roots

<sup>a</sup> Family.

EN28 from inflorescence. This apparent inconsistencies could be due to low levels of expression. For example, it was possible to amplify a very low intensity band corresponding to gene EN28 in flowers and young siliques using higher quantities of cDNA template and a high number of PCR cycles (data not shown).

#### 4. Discussion

We identified a total of 509 ANK repeats in the *Arabidopsis* proteins coded by 105 genes, which represent 0.4% of the total *Arabidopsis* genes. This number is higher to a previous estimation of 0.25%, but is similar to the percentages estimated for humans, *Drosophila* and *C. elegans* (Jebanathirajah et al., 2002). Few of the amino acids in the ANK repeats are well conserved (Table 3). The use of more precise criteria to identify ANK repeats allow us to recognize many previously nonannotated repeats. Evaluating sequence conservation it becomes obvious that the terminal repeats in the arrays deviate more from the general consensus than those located centrally and the same was observed in animal proteins (Bork, 1993).

Although few of the 33 amino acids that compose the ANK repeats are conserved, several strictly conserved hydrophobic positions can be observed either in *Arabidopsis* and animals. This conservation is necessary to maintain the secondary structure that is essential for its function in protein–protein interactions (Bork, 1993; Rhode and Bork, 1993). Few plant proteins containing ANK repeats have been characterized and any experimental data demonstrates the role of the ANK repeats in plants. However, the conservation of the strictly hydrophobic positions in *Arabidopsis* suggests that ANK repeats may have similar functions in plants and animals. Many plant proteins containing ANK repeats are multidomain molecules in which ANK repeats are combined with other unrelated structural modules. The presence of ANK repeats on so diverse proteins makes a common function such as an enzymatic activity extremely unlikely and supports the idea that ANK repeats are involved in mediate protein–protein interactions also in plants.

The most abundant group of structurally similar ANK repeat-containing proteins counts for 40 elements: 37 proteins containing ANK repeats and transmembrane domains (Fig. 4), and three more that may represent truncated forms containing only the N-terminal ANK repeats. The 40 *AnkTm* genes are divided into six families. The evolution of these gene families has been complex. *AtAnkTm* genes are distributed all over the genome but not uniformly. This type of distribution seems to be common to other gene families and, for example, is similar to the distribution of the bHLH genes, although the regions of high and low density are not the same in both cases (Toledo-Ortiz et al., 2003). Inter- and intrachromosome and tandem array duplications have certainly played a role in amplify the number of *AtAnkTm* genes.

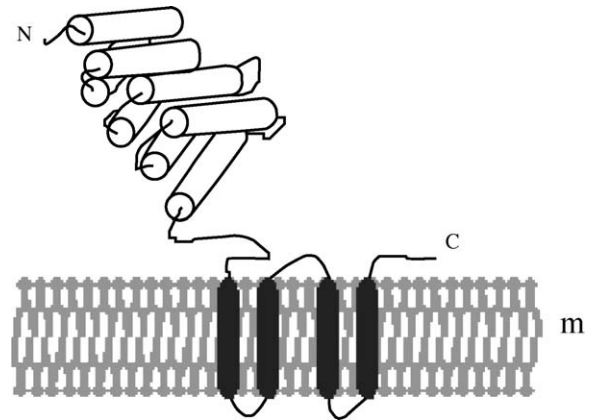


Fig. 4. Schematic representation of the *Arabidopsis* ANKTM proteins. Hypothetic representation of an ANKTM protein and the associated membrane (m). N and C ends of the protein are indicated by letters. Grey cylinders are the ankyrin repeats and black cylinders the transmembrane domains.

Sixteen of the *AtAnkTm* genes are located in different tandem arrays. Sets of genes organized in tandem arrays are very common in *Arabidopsis*.

The interaction specificity of the ANK repeats seems to be determined by its amino acid sequence (Bennett, 1992). When comparing *AtANKTM* families, the sequences of the ANK repeats are not well conserved between families but highly conserved within the same family. This conservation suggests that *AtANKTM* proteins of the same family may interact with the same or similar proteins. The variability in the expression patterns of genes of the same family indicates that although they could interact with similar proteins, their roles may not be redundant. Some of the *AtAnkTm* genes are widely expressed suggesting that they may have pleiotropic or general functions, whereas others have a more restricted expression and perhaps more specific functions. Only one of the *AtAnkTm* genes have been studied previously (*ACD6*, *EN11*) and codes a protein involved in salicylic acid signaling in defense responses (Lu et al., 2003). A mutant of this gene shows spontaneous cell-death and increased disease resistance. The function of this protein at the molecular level remains unknown. Proteins with similar domain organization have been found in other plant species (monocot and dicot) but their functions also remain unknown. Similar domain organization have been found also in some animal proteins whose functions are known and can suggest some roles for the *AtAnkTm* genes: (a) Membrane receptors, as ANKTM1, a menthol- and cold-activated channel (Story et al., 2003) and the human vanilloid receptor; (b) membrane channels, as the human CaT1 and CaT2 calcium entry channels (Peng et al., 2001), the OTRPC4 cation channel or the transient receptor potential channel present in taste receptor cells; or (c) membrane anchorage proteins that attach other proteins to the membrane in a similar role as ankyrin protein does (Bennett, 1992). Any of the three possibilities is compatible with the function of *ACD6* (Lu et al., 2003).



Proteins of *AtAnkTm* family 2 have certain sequence similarity with the maize TM20 protein in the region of the transmembrane domains. TM20 is a protein necessary for normal embryo development and contains twenty hydrophobic segments that can be grouped in five repeats formed by four segments (Stiefel et al., 1999). A possible function of TM20 is to act as an auxin membrane transporter (T Jahrmann, personal communication). In *Arabidopsis*, no gene coding for a protein with 20 transmembrane domain is present. A possibility could be that in *Arabidopsis* the function of the maize TM20 protein is carried out by a complex of AtANKTM proteins bound by the ANK repeats.

Mutant analysis, double-hybrid assays and cell localization experiments will give us the necessary clues to understand the functions and molecular interactions of AtANKTM proteins.

## Acknowledgements

C.B. was the recipient of a fellowship of the Universitat Autònoma de Barcelona - Fundació President Allende. C.M.V. is recipient of a “Ramon y Cajal” contract from the Spanish Ministry of Science. This work has been carried out thanks to the grant BIO2001-1721 from Plan Nacional de Investigación Científica y Técnica, to a grant from programme MAZE (European Union) and was done within the framework of Centre de Referència de Biotecnologia de la Generalitat de Catalunya.

## References

- Albert, S., Despres, B., Guilleminot, J., Bechtold, N., Pelletier, G., Delseny, M., Devic, M., 1999. The EMB 506 gene encodes a novel ankyrin repeat containing protein that is essential for the normal development of *Arabidopsis* embryos. *Plant J.* 17, 169–179.
- Bardwell, V.J., Treisman, R., 1994. The POZ domain: a conserved protein–protein interaction motif. *Genes Dev.* 8, 1664–1677.
- Bennett, V., 1992. Ankyrins: adaptors between diverse plasma membrane proteins and the cytoplasm. *J. Biol. Chem.* 267, 8703–8706.
- Bork, P., 1993. Hundreds of ankyrin-like repeats in functionally diverse proteins: mobile modules that cross phyla horizontally? *Proteins* 17, 363–374.
- Bouché, N., Scharlat, A., Snedden, W., Bouchez, D., Fromm, H., 2002. A Novel family of calmodulin-binding transcription activators in multicellular organisms. *J. Biol. Chem.* 277, 21851–21861.
- Cao, H., Glazebrook, J., Clarke, J.D., Volko, S., Dong, X., 1997. The *Arabidopsis* NPR1 gene that controls systemic acquired resistance encodes a novel protein containing ankyrin repeats. *Cell* 88, 57–63.
- Chinchilla, D., Merchan, F., Megias, M., Kondorosi, A., Sousa, C., Crespi, M., 2003. Ankyrin protein kinases: a novel type of plant kinase gene whose expression is induced by osmotic stress in alfalfa. *Plant Mol. Biol.* 51, 555–566.
- Chye, M.L., Li, H.Y., Yung, M.H., 2000. Single amino acid substitutions at the acyl-CoA-binding domain interrupt <sup>14</sup>C]palmitoyl-CoA binding of ACBP2, an *Arabidopsis* acyl-CoA-binding protein with ankyrin repeats. *Plant Mol. Biol.* 44, 711–721.
- da Costa e Silva, O., 1994. CG-1, a parsley light-induced DNA-binding protein. *Plant Mol. Biol.* 25, 921–924.
- Isono, K., Yamamoto, H., Satoh, K., Kobayashi, H., 1999. An *Arabidopsis* cDNA encoding a DNA-binding protein that is highly similar to the DEAH family of RNA/DNA helicase genes. *Nucleic Acids Res.* 27, 3728–3735.
- Jebanathirajah, J.A., Peri, S., Pandley, A., 2002. Toll and interleukin-1 receptor (TIR) domain-containing proteins in plants: a genomic perspective. *Trends Plant Sci.* 7, 388–391.
- Klimyuk, V.I., Persello-Cartieaux, F., Havaux, M., Contard, P., Schuenemann, D., Meierhoff, K., Gouet, P., Jones, J.D.G., Hoffman, N.E., Nussaume, L., 1999. A chloro protein encoded by the *Arabidopsis* gene CAO is a plant specific component of the chloroplast signal recognition particle. *Plant Cell* 11, 87–99.
- Lin, J.H., Makris, A., McMahon, C., Bear, S.E., Patriotic, C., Prasad, V.R., Brent, R., Golemis, E.A., Tschilis, P.N., 1999. The ankyrin repeat-containing adaptor protein Tvl-1 is a novel substrate and regulator of Raf-1. *J. Biol. Chem.* 274, 14706–14715.
- Lu, H., Rate, D.N., Song, J.T., Greenberg, J.T., 2003. ACD6, a novel ankyrin protein, is a regulator and an effector of salicylic acid signaling in the *Arabidopsis* defense response. *Plant Cell* 15, 2408–2420.
- Michaely, P., Bennett, V., 1992. The ANK repeat: a ubiquitous motif involved in macromolecular recognition. *Trends Cell Biol.* 2, 127–129.
- Mosavi, L.K., Minor Jr., D.L., Peng, Z.Y., 2002. Consensus-derived structural determinants of the ankyrin repeat motif. *Proc. Natl. Acad. Sci.* 99, 16029–16034.
- Peck, S.C., Nühse, T.S., Hess, D., Iglesias, A., Mein, F., Boller, T., 2001. Directed proteomics identifies a plant-specific protein rapidly phosphorylated in response to bacterial and fungal elicitors. *Plant Cell* 13, 1467–1475.
- Peng, J.B., Brown, E.M., Hediger, M.A., 2001. Structural conservation of the genes encoding CaT1, CaT2, and related cation channels. *Genomics* 76, 99–109.
- Pilot, G., Pratelli, R., Gaymard, F., Meyer, Y., Sentenac, H., 2003. Five-Group distribution of the shaker-like K<sup>+</sup> channel family in high plants. *J. Mol. Evol.* 56, 418–434.
- Rhode, K., Bork, P., 1993. A fast, sensitive pattern-matching approach for protein sequences. *CABIOS* 9, 183–189.
- Sedgwick, S.G., Smerdon, S.J., 1999. The ankyrin repeat: a diversity of interactions on a common structural framework. *Trends Biochem. Sci.* 24, 311–316.
- Sentenac, H., Bonneaud, N., Minet, M., Lacroute, F., Salmon, J.M., Gaymard, F., Grignon, C., 1992. Cloning and expression in yeast of a plant potassium ion transport system. *Science* 256, 663–665.
- Stiefel, V., Becerra, E.L., Roca, R., Bastida, M., Jahrmann, T., Graziano, E., Puigdomenech, P., 1999. TM20, a gene coding for a new class of transmembrane proteins expressed in the meristematic tissues of maize. *J. Biol. Chem.* 274, 27734–27739.
- Story, G.M., Peier, A.M., Reeve, A.J., Eid, S.R., Mosbacher, J., Hricik, T.R., Earley, T.J., Hergarden, A.C., Andersson, D.A., Hwang, S.W., McIntyre, P., Jegla, T., Bevan, S., Patapoutian, A., 2003. ANKTM1, a TRP-like channel expressed in nociceptive neurons, is activated by cold temperatures. *Cell* 112, 819–829.
- Toledo-Ortiz, G., Huq, E., Quail, P.H., 2003. The *Arabidopsis* basic/helix-loop-helix transcription factor family. *Plant Cell* 15, 1749–1770.
- Unoki, M., Nakamura, Y., 2001. Growth-suppressive effects of BPOZ and EGR2, two genes involved in the PTEN signaling pathway. *Oncogene* 20, 4457–4465.
- Vicient, C.M., Delseny, M., 1999. Isolation of total RNA from *Arabidopsis thaliana* seeds. *Anal. Biochem.* 268, 412–413.
- Vicient, C.M., Hull, G., Guilleminot, J., Devic, M., Delseny, M., 2000. Differential expression of the *Arabidopsis* genes coding for Em-like proteins. *J. Exp. Bot.* 51, 1211–1220.
- Yan, J., Wang, J., Zhang, H., 2002. An ankyrin repeat-containing protein plays a role in both disease resistance and antioxidation metabolism. *Plant J.* 29, 193–202.
- Zhang, H., Scheirer, D.C., Fowle, W., Goodman, H.M., 1992. Expression of antisense or sense RNA of an ankyrin repeat-containing gene blocks chloroplast differentiation in *Arabidopsis*. *Plant Cell* 4, 1575–1588.